FROM THE EDITOR

This issue is mainly devoted to household surveys in transition countries. Particularly, two main types of surveys are considered: Household Budget Survey (HBS), and Labour Force Survey (LFS). The following features of the surveys are discussed: sampling frame, sample design, size of samples, methods of estimation, estimation of sampling errors, nonresponse rates, survey costs, and design effects. The HBS has had already some tradition in transition countries. However, redesign of the HBS was necessary in each country. LFS is a completely new survey and has been introduced in the last decade or is to be introduced soon. The articles from the following countries are presented: the Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Romania and Slovenia.

There are also two articles in the section ***Other Articles***.

The articles on household surveys in some transition countries include:

1. ***Household Surveys in Hungary*** *(by* Ö. Éltetõ and L. Mihályffy from Hungary). The paper describes primarily the main methodological characteristics of two continuous household surveys conducted by the Hungarian CSO: the Household Budget Survey and the Labour Force Survey. Issues regarding the sampling design, rotation system, conducting the surveys, response rates, sampling and non-sampling errors, treating of nonresponse, design effect, method of calibration, survey cost, corrections and imputations are discussed. Plans on modifying the sampling designs of the surveys on the basis of the 2001 Population and Housing Census are also mentioned. Besides, some methodological issues of a few other household surveys carried out in recent years like the 2 percent Microcensus in 1996, the Housing Survey in 1999 and the Time Use Survey in 1999-2000 are also presented.
2. ***The Labour Force Survey in the Czech Republic*** (by M. Kopecky, I. Makalous and Z. Pavlickova from the Czech Republic). The paper presents some aspects of designing and implementation of the Labour Force Survey in the Czech Republic, which started in 1992 as a continuous quarterly survey. The sample was selected in two stages and stratified by regions. Some features of the survey are described: sample design, weighting techniques, non-response rates, cost components of the survey, and other organisational aspects.
3. ***The Household Sample Surveys in Poland*** (J. Kordos, B. Lednicki and M. Zyra from Poland). The authors present the design and implementation of the household sample surveys carried out by the Central Statistical Office of

Poland (GUS) in the last decade. I.e. during the transition period. The paper starts with a general description of household surveys carried out by the GUS before the transition period, followed by the design and implementation of the continuous household surveys, i.e. the household budget surveys and the labour force survey, and ad hoc surveys, such as living conditions surveys, health status surveys, demographic surveys, and other social surveys. Special attention is paid to sample designs, estimation methods, sampling errors, design effects, costs of the surveys, nonresponse rates, and dissemination of the results. Future improvement of household surveys after the Population Census 2002 is stressed.

4. ***The Household Sample Surveys in Lithuania*** (by D. Krapavickaitë from Lithuania). The current state of two main household surveys in Lithuania – the labour force survey and household budget survey – is discussed. Sampling designs, comments on nonresponse and some main estimates with comments are presented. Other household surveys are also reviewed.

5. ***The Estonian Household Sample Surveys – Focus on the Labour Force Survey*** (by M. Kurvits, K. Sõstra and I. Traat from Estonia). The authors present development and current face of the Estonian Labour Force Survey. The specific features of data collection programme, the sampling design and its changes, the nonresponse and weighting procedure are described. Some comparisons with the Household Budget Survey are made.

6. ***Household Surveys in Latvia*** (by J. Lapins, E. Vaskis, Z. Priede, and S. Balina from Latvia). This paper presents a detailed analysis of the methodological and organisational principles and solutions applied to two of the surveys, the Household Budget Survey and the Labour Force Survey. The main changes that have already been made (or planned to be made) in the redesigning process of these surveys in the course of 2001 and 2002 are also discussed.

7. ***Household Surveys – Main Tools of Social Statistics*** (by F. Panduru, A. Vasile and S. Pisica from Romania). This paper presents the main characteristics of sampling surveys of households, implemented in the social statistics system in Romania during the transition period to the market economy, focused on Household Budget Survey and Labour Force Survey. Problems referring to the surveys methodological out-look are presented, with a special accent on sampling aspects (sampling design, sample size, method of estimation), non-response rates, surveys cost, harmonisation with EU standards.

8. ***Household Sample Surveys in Slovenia*** (by V. Vehovar, M Zaletel, T. Novak, M. Arnez, K. Rutar from Slovenia) The paper describes the sample design and implementation of two major household surveys conducted by the Statistical Office of the Republic of Slovenia: Labour Force Survey and Household

Budget Survey. Both surveys were redesigned in 1997. The new sample designs have basically followed the Eurostat guidelines, however, they incorporate certain country-specific features, too. The paper also provides the background information about 10-year transition process of the Slovenian Statistical Office from a regional statistical bureau – performing almost no sample surveys – to a modern European statistical agency. Further on, the sample designs are described, together with the discussion of open issues. Response rates, costs, sampling and non-sampling errors are discussed as well as the specifics of the corresponding survey modes and patterns of panel rotation.

There are two articles in the section ***Other Articles****:*

9. ***An Alternative Measure of Unemployment in Transition Economies – Idea of Free-Market Unemployment*** (G. Kuczynski and K. Strzala from Poland). Economies in transition are subject to privatisation process that may be perceived as a permanent structural shock to the economy and the traditional measures of unemployment seem to be inadequate. That is why the authors suggest decomposing the observed unemployment rate into two fractions – privatisation and free market unemployment. The former they consider as the level of unemployment directly subject to the transformation process, the latter – as unemployment which formation is subject only to the market forces. In order to reach this aim the Blanchard-Quah decomposition is used, which allows for direct derivation of the privatisation unemployment. The free market unemployment is derived as the difference between the observed and the privatisation unemployment.

10. ***Empirical Studies of Generalized Classes of Ratio and Product Type Estimators under a Linear Model*** **(**by G.N. Singh from India). In this work, two generalized classes of ratio and product type estimators have been considered for estimating the population mean. Their performances have been examined under a linear model and gamma distributed auxiliary variable. Empirical comparisons are made with sample mean, ratio and product estimators and the recommendations made.

Jan Kordos
The Editor

STATISTICS IN TRANSITION, June 2002
Vol. 5, No. 4, pp. 521–540

HOUSEHOLD SURVEYS IN HUNGARY

Ödön Éltető and László Mihályffy[1]

Abstract

The paper describes primarily the main methodological characteristics of two continuous household surveys conducted by the HCSO, the Household Budget Survey (HBS) and the Labour Force Survey (LFS). Issues regarding the sampling design, rotation system, conducting the surveys, response rates, sampling and non-sampling errors, treating of nonresponse, design effect, method of calibration, survey cost, corrections and imputations are discussed. Plans on modifying the sampling designs of the surveys on the basis of the 2001 Population and Housing Census are also mentioned. Besides, mention is made of some methodological issues of a few other household surveys carried out in recent years like the 2 percent Microcensus in 1996, the Housing Survey in 1999 and the Time Use Survey in 1999-2000.

1. Introduction

Household surveys have a rather long tradition in Hungary. The HBS began as early as the 1950s, based first on quota samples, then, after developing the so-called Unified System of Household Surveys (USHS) in the mid 1970s, on probability samples. The quinquennial Income Surveys introduced in 1963 were based from the beginning on probability samples. There were a number of household surveys carried out within the frame of the USHS, especially in the 80s, as e.g. a Time Use Survey, a Prestige Survey (prestige of the various occupations), a survey on Living Conditions and Social Stratification, etc.

The LFS was introduced in the transition period, in 1992; its methodology was harmonised with ILO notions and definitions. The HBS, which was a biennial survey in the 1980s, was made continuous again since 1993. It should be mentioned that the four 25 percent subsamples of the 2 percent Microcensus in 1996 were used to carry out important household surveys partly simultaneously (income survey, wealth survey and survey on tourism), partly later (e.g. housing survey).

[1] Central Statistical Office, P.O. Box 51, H-1525 Budapest, Hungary – odon.elteto(%)office.ksh.hu and Laszlo.mihalyffy(%)office.ksh.hu

The HCSO has a directorate in every county (including the capital). The fieldwork of all household surveys is directed and supervised by the staff of these county statistical directorates. There is a more or less stable network of interviewers, who previously used to be employees of the HCSO, but from the mid 1990s they work on a contract basis.

Data capture is generally done at the county directorates, then data are transferred through the network to the Department of Informatics of the Head Office in Budapest, where the processing is done.The paper proceeds as follows: we begin with a description of the main common features of the sampling designs of the two continuous household surveys, the HBS and the LFS. Then the theoretical and practical aspects of the weighting and calibration procedures used in the HBS and the LFS are discussed. Section 4 presents the method and certain results of sampling error computations concerning the HBS and the LFS. The next two sections deal with special features of the HBS and the LFS, respectively. Then a few other household surveys are discussed shortly. The paper ends with e few concluding remarks.

2. Common Features of the Sampling Designs for the HBS and the LFS

As the samples of both the HBS and the LFS are subsamples selected from a common sampling frame, the USHS sample, no wonder that the sampling designs of the two surveys are very similar in many respects. For both surveys the samples have been selected in three stages, except the larger self-representing cities, i.e. cities with more than 20 thousand and 15 thousand inhabitants, respectively, where the selection has been carried out in two stages. In the case of non-self-representing localities, the primary sampling units (PSUs) are the localities themselves, the secondary units (SSUs) are the census enumeration districts (EDs) and the ultimate sampling units are the dwellings. In self-representing cities the EDs are the PSUs.

Both the localities and the EDs are selected with probability proportional to size (measured by the number of dwellings), while dwellings within EDs with equal probability. Accordingly, within each size stratum of PSUs an approximately equal number of SSUs is selected and within a stratum the sample contains from each ED an equal number of dwellings. This sampling procedure ensures that within a given stratum each household has the same probability of being selected. Before selection, localities have been stratified into eight and seven size categories, respectively, and as the selection has been carried out by counties, this involves an implicit geographic stratification as well. The sample was not proportionately allocated to the strata, the sampling rate was higher in the case of smaller localities than in larger cities and especially in Budapest. Besides, EDs in remote, sparsely inhabited places of the localities have been treated as belonging to separate strata.

Ever since 1993, the HBS and LFS samples have been disjoint, but the nature of distinction has changed over time. From 1993 to 1995, the two samples had no dwellings in common, but they shared a number of EDs which implied that in terms of EDs the HBS sample was a subsample of the LFS sample. This involved some problems, thus a stronger distinction seemed necessary. Since 1996, the two samples have no longer had EDs and even small localities in common, provided the size of the latter did not exceed 5000 inhabitants.

3. Costs of the Surveys

Both the HBS and the LFS are rather expensive surveys. In 2000 the direct expenditures on the survey, i.e. remuneration of the interviewers, incentives for co-operating households and material costs amounted to 84,769 thousand HUF, corresponding that time roughly to €326,000. The above amount naturally does not contain the salaries and other wages like incomes of the personnel in the central and county offices of the HCSO managing the survey nor the net costs of the publications presenting the results of the survey. Table 1 below shows the structure of the direct expenditures on HBS in 2000 both in absolute amounts (in thousands of HUF) and in percentages.

Table 1. Cost Structure of the HBS in the Year 2000

Cost Components	Costs in	
	1000 HUF	Percentages
Costs Connected with Monthly Diaries	38,638	45.6
– End of Year Questionnaires	9,346	11.0
– Call on Non-responding Households	1,136	1.3
Incentives to Co-operating Households	19,751	23.3
Premium to Inter-viewers	4,841	5.7
Material Costs	11,057	13.1
Total	84,769	100.0

Expenditures on the LFS exceed those on the HBS. In 2000 direct expenditures on the survey amounted to 112,496 thousand HUF corresponding to €432,000. However, considering that a household is called four times a year and no incentive is given to the co-operating households, the expenditures per household are considerably lower than those of the HBS. Table 2. below shows the structure of costs of the LFS also both in absolute terms (in thousand HUF) and in percentages.

Table 2. Cost Structure of the LFS in the Year 2000

Cost Components	Costs in	
	1000 HUF	Percentages
Call on Addresses	5,767	5.1
Costs Connected with Household Questionnaires	17,030	15.1
– Activity Questionnaires	55,251	49.1
– Supplementary Questionnaires	11,001	9.8
Premium to Interviewers	8,774	7.8
Material Costs	14,673	13.1
Total	112,496	100.0

4. Weighting and Calibration

4.1. Design Weights

The household surveys of the HCSO have a common philosophy, which is manifested first of all in the structure of the samples, definition of the sample weights and the methods of estimation. The surveys are all based on samples of dwellings, and the design ensures the conditions of using the familiar Horvitz-Thompson estimator, i.e. that estimated totals are weighted sums of the observations and the *design* weights are reciprocals of the inclusion probabilities.

In any household survey of the HCSO, there is a single design weight for each stratum defined by the design. Provided there are H strata in the sample, the design weight is defined for stratum h basically as follows:

$$d = d_h = L_h / \ell_h$$

where L_h is the total number of dwellings in the stratum of which ℓ_h dwellings are included in the sample. Note that, as of present, $H = 98$ in the case of the HBS and $H = 130$ in the case of the LFS. Recently, this definition has slightly been modified so that numbers of non-vacant dwellings only occur both in the numerator and denominator of the ratio, moreover, the denominator is restricted only to those dwellings in the sample which provided at least one completed interview. This modification compensates to some extent for nonresponse at household level.

4.2. Calibration

In today's practice of household surveys, calibration is extensively used to improve survey estimates. If properly used, it can compensate for nonresponse and also for deficiencies of the coverage. Calibration was introduced in the HBS in 1994 to reconcile adjusted estimates of households and those of individuals. Since 1995, the same approach has been in use for the LFS data. In what follows

the current calibration technique used in the household surveys of the HCSO is described.

4.2.1. The Mathematical Model of Calibration

Based on what was said in para. 4.1, totals in a household survey of the HCSO can be estimated in the form

$$\hat{Y} = \sum_{j=1}^{n} d_j y_j \tag{1}$$

where $y_1, y_2, ..., y_n$ are observed values of the study variable on the units of the sample and d_j is the design weight for unit *j*. From the form of this estimator it follows immediately that any instance of nonresponse involves a downward bias in the value of $\hat{Y}$. In such a case, which often occurs in practice, one would try to modify the weights of the *responding* units to compensate somehow for the issing information. Calibration is a device to manage this compensation in a suitable way. In a household survey there are usually some variables signalling the effect of missing responses; if e.g. y_j in (1) is the number of males aged 20-24 in household j, and the population total *Y* for this category is known, the difference $Y-\hat{Y}$ gives some information on the bias due to nonresponse. In this case it is easy to modify the design weights d_j so that the bias be eliminated, and the variable y_j is called an *auxiliary* or *control* variable. In the general case, there are more control variables; to distinguish them from the study variables, they are usually denoted by $x_1, x_2, ..., x_m$, respectively. Their population totals should be known and are denoted by $X_1, X_2, ..., X_m$, respectively. In view of (1), it is straightforward to require that the equations

$$X_1 = \sum_{j=1}^{n} w_j x_{1j}, \quad X_2 = \sum_{j=1}^{n} w_j x_{2j}, \quad ..., \quad X_m = \sum_{j=1}^{n} w_j x_{mj} \tag{2}$$

may hold. In these *calibration equations*, *n* stands for the size of the *responding* part of the sample, and the *calibrated* or *final* weights w_j replace the design weights d_j. Since the calibration modifies the selection probabilities, it is important that the calibrated weights may be in some sense close to the design weights. The distance between the two kinds of weights is measured by a *distance function*

$$F(d_1, d_2, ..., d_n, w_1, w_2, ..., w_n), \tag{3}$$

which is supposed to be convex and continuously differentiable. The *calibration problem* is then stated as follows: minimise the distance function *F* subject to the

constraints (2). Sometimes additional constraints are also considered in the calibration problem, e.g.

$$0 < L \leq \ w_j / d_j \leq U, \quad j = 1, 2, ..., n$$

where L and U are suitable lower and upper bounds for the relative change in the weights. The most common choice of the distance function is either

$$F_1 = \sum_{j=1}^{n} \frac{(w_j - d_j)^2}{d_j} \qquad \text{or} \qquad F_2 = \sum_{j=1}^{n} (w_j \log \frac{w_j}{d_j} - w_j + d_j) .$$

Choosing F_1 implies that sample estimates of totals of study variables will be generalised regression estimates. In this case the calibrated weights as well as the variance of the estimates can be computed with analytic expressions. If F_2 is used – which is called information divergence function, – the calibrated weights are computed with an iterative but very simple algorithm. It runs as follows.
Start with the equations (2) and set the *initial* values of the w_j s equal to the d_j s. As the equations will not hold, assign a multiplier r to each of the m relations to restore equality. This yields m different updates to each weight w_j, say

$$r_1 w_j, \ r_2 w_j, \ ..., \ r_m w_j .$$

To get a single update, take the average of these values weighting them with x_{1j}, x_{2j}, ..., x_{mj}, and go back to the previous step: find new multipliers r_1', r_2', ..., r_m' to restore equality, then average the new updates again, etc. Repeat until convergence is found.

In the Hungarian household surveys this second method of calibration called generalised iterative scaling or just raking has been adopted. There is a broad literature on calibration; the basic facts can be found in Darroch and Ratcliff (1972), Deville and Särndal (1992), and Chambers and Skinner (1999).

4.2.2. Auxiliary Variables Used for Calibrating the HBS Data

The current calibration strategy used in the HBS is the result of gradual improvements starting at a point when it was realised that calibrating on demographic profiles alone did not produce satisfactory results. The set of auxiliary variables used to calibrate the HBS is as follows:

- sex and age (2×4 categories),
- economic activity (9 categories),
- level of education (3 categories),
- household type (3 categories).

Two kinds of geographic breakdowns are used when calibrating HBS data. The main areas are defined as follows:

- the capital city (Budapest),
- the major cities (county towns plus four cities with 50,000 or more inhabitants),
- the rest of the country.

The NUTS II level regions are defined in accordance with Eurostat recommendations. There are seven of them, with population ranging from 700,000 to 3 million.

Table 3. Auxiliary Variables, Controls and Calibrated Estimates in the Hungarian HBS Annual Data, 1999

Area Unit, Auxiliary Variable	Estimate	Control	Rel.Difference, %
West Trans-Danubia, Male, 0-14	83835	83848	-0.02
" " 15-29	116372	116367	0.00
" " 30-59	198915	198927	-0.01
" " 60 and older	74476	74462	0.02
Northern Hungary, Female, 0-14	114426	114430	-0.00
" " 15-29	133726	133739	-0.01
" " 30-59	256658	256676	0.01
" " 60 and older	154752	154783	-0.02
Budapest, Employed	601574	601586	-0.00
" Self-Employed	116900	116907	-0.01
" Unemployed[a]	32956	32957	-0.00
" Unemployed[b]	16053	16053	0.00
" Child Care Receiver	43721	43717	0.01
" Pensioner	529169	529167	0.00
" School Drop-Out, under 14	113655	113649	0.01
" Schoolchild, Student	312336	312319	0.01
" Other Dependant	46180	46181	-0.00
Other Localities, Level of Education: 1[c].	1270292	1270290	0.00
" Level of Education: 2[c]	630402	630405	-0.00
" Level of Education: 3[c].	238343	238350	-0.00
" Household, Total	2243202	2242554	0.03
" One-Person Households	477870	478000	-0.03
" Households with Child	835979	836000	-0.00

a) Unemployed receiving benefit, *b*) Unemployed receiving no benefit, *c*) Active earners only

When processing the quarterly data of the HBS, calibration is performed for the three main areas separately. In the case of annual data, age-sex categories are crossed with the NUTS II level regions, while the other categories with the three

main areas. This implies that the calibration problem cannot be decomposed in a set of problems of smaller size; e.g. in 1999 we had a sizeable problem of calibrating 10185 weights subject to 101 calibration equations. A part of the results indicating the accuracy of calibration is shown in Table 3.

4.2.3. Auxiliary Variables Used for Calibrating the LFS Data

In the LFS, only demographic calibration is used, which represents a much simpler case than that in the HBS. On the basis of monthly data collection, calibration is applied to the weights of the monthly subsample, and – since there is no overlap between the subsamples of consecutive months – calibrated quarterly weights are derived from the monthly weights simply by dividing them by 3.

For the LFS in any month, calibration is done separately for the capital and for the 19 counties. For each of these area units, the following controls are used:

- totals of age-sex groups defined by the ages 0-14, 5-19, 20-24, 25-29, 30-39, 40-54, 55-59, 60-69, 71-74, 75 or more;
- total resident population in the major cities in the county;
- total number of households.

Since the monthly subsample of the LFS contains 11,000-12,000 responding households, and even the subsamples of the area units are large enough, the method of generalised raking works fairly well for the LFS. The pattern of deviations of estimated totals from the corresponding controls is similar to that displayed in Table 3.

4.2.4. Calibration from the Aspects of Computation and Quality

The computer programs used for calibrating the estimates of the household surveys of the HCSO are based on generalised raking discussed in subsection 3.2.1; they were prepared by HCSO methodologists in the Interactive Matrix Language (IML) of the SAS program package. The method of generalised raking always produces a *convergent* sequence of *approximate* solutions of the calibration equations (2); and the limit of that sequence is not a *feasible* solution of (2) only if no such solution exists. As it is illustrated by Table 3, the accuracy of calibration is satisfactory in the applications considered. Most calibration problems can be solved separately for the underlying area units, which means that the original problem can be replaced by a set of similar problems of smaller size. If this is the case, the complexity of the problem and the CPU time needed to carry out calibration are considerably reduced.

Two further aspects in connection with calibration relate to the profile of calibrated weights and the number of observations in the cross-classes or cells for which the *weighted* number of observations is controlled by the procedure. In the annual HBS in 1999 there were all in all 92 cells of individuals of which only one had less than 20 observations. As for the ranges of calibrated sample weights, the minimal and maximal values can be regarded as outliers; e.g. out of 10,000-

11,000 weights in the monthly LFS, on the average only 100 exceed the value 1000. It is important to note that even before calibration, the weights spread in a rather wide range.

5. Sampling Error, Design Effect

5.1. Sampling Error Computation

Analytical variance expressions are not available for estimating the sampling error of calibrated estimates, therefore some approximate method should be used. In processing the household surveys of the HCSO, the jackknife method has been adopted for this purpose. Though this technique is widely known, it might be useful to recall its definition in the case of multi-stage samples.

Table 4. Some Data of the LFS in the First Quarter of 2001 and the Corresponding Sampling Errors (s.e.) at Probability Level 95 %

Age/Sex Groups	Emplo-yed	Un-Emplo-yed	In Labour Force	Out of Labour Force	Total	Rate of Partici-pation	Rate of Unem-ployment
15-74	3836231	245528	4081759	3581936	7663695	53.3	6.0
s.e.	40556	15021	40556	40556	..	0.5	0.4
15-19	46831	14007	60838	554740	615578	9.9	23.0
s.e.	4705	2654	5429	5429	..	0.9	3.9
20-24	418598	46613	465211	343948	809159	57.5	10.0
s.e.	12370	5551	12529	12529	..	1.5	1.1
25-29	559373	40552	599925	190295	790220	75.9	6.8
s.e.	10842	5266	10067	10067	..	1.3	0.9
30-39	967834	59018	1026852	254207	1281059	80.2	5.8
s.e.	13057	6026	12052	12052	..	0.9	0.6
40-54	1567980	76871	1644851	528614	2173465	75.7	4.7
s.e.	18318	6816	17466	17466	..	0.8	0.4
55-59	215842	7079	222921	388350	611271	36.5	3.2
s.e.	9585	2037	9705	9705	..	1.6	0.9
60-69	54214	1388	55602	912854	968456	5.7	2.5
s.e.	6074	759	6074	6074	..	0.6	1.5
70-74	5559	0	5559	408928	414487	1.3	0.0
s.e.	1625	0	1625	1625	..	0.4	0.0
Male	2104835	152664	2257499	1411859	3669358	61.5	6.8
s.e.	25172	10788	24453	24453	..	0.7	0.5
Female	1731396	92864	1824260	2170077	3994337	45.7	5.1
s.e.	25052	7829	25052	25052	..	0.6	0.4

Consider first a sample without stratification. Suppose there are k primary sampling units (PSUs) in the sample selected with PPS, and $\hat{Y}$ is the sample

estimate of some parameter Y of the population; no matter if Y is a total, a mean or a proportion. Denote $\hat{Y}(k)$ the so-called pseudo-estimate obtained by omitting the kth PSU from the sample and estimating Y on the basis of the remaining k-1 primary sampling units. $\hat{Y}(k)$ is computed by nonresponse adjustment, just as in the case where the observations in the kth PSU are missing owing to nonresponse. The jackknife estimate of the variance of $\hat{Y}$ is as follows:

$$\mathrm{var}(\hat{Y}) = \frac{k-1}{k} \sum_{l=1}^{k} (\hat{Y}(l) - \hat{Y})^2$$

In the case of stratified samples, the pseudo-estimates are similarly defined as in the previous case, but when omitting the kth PSU, sample weights are modified only in the stratum containing the omitted PSU.

The jackknife method is regularly used to estimate the sampling error in the LFS, both for quarterly and monthly data, though the latter are not published. Up to now, sampling error computations have been rather occasional in the HBS, but it is planned to make them regular in the future. The jackknife technique will probably be replaced by the bootstrap method for that survey. Sampling error computations based on the jackknife method were carried out for a number of data of the 1996 Microcensus, too. Some LFS data together with their sampling error are shown in Table 4

The LFS complies with the recommendation of the Eurostat as stated in the Council Regulation (EC) No. 577/98, 1998. According to that recommendation, the coefficient of variation of the estimated level of the unemployed should not exceed 8 % if it is measured for a NUTS II region, and the unemployed there amount to at least 5 % of the working age population. At national level, the standard error of the change in the level of the unemployed between two consecutive quarters should not exceed 3 % of the level.

5.2. Design Effect

The design effect introduced by L. Kish (1910-2000) is an indicator measuring the efficiency of a probability sample. For a sample estimate $\hat{Y}$, it is defined as

$$deft^2 = \mathrm{Var}(\hat{Y}) / \mathrm{Var}_o(\hat{Y})$$

where $\mathrm{Var}(\hat{Y})$ is the variance of the estimate by the design of the sample and $\mathrm{Var}_o(\hat{Y})$ is the variance which would have been obtained if the sample had been a simple random sample having the same sample size as the working sample in consideration. Given some complex design, the $deft^2$ statistics assumes different values for different estimates from the same sample, or for data of the same characteristic estimated from different subsamples. To get an overall picture of the sample and the estimator associated to it, it is useful to compute the $deft^2$ statistics for different characteristics and different breakdowns. In Table 5, the

positive square root of $deft^2$ is given for some estimates of the Hungarian HBS and the LFS. For estimating $Var_o(\hat{Y})$, the variance of the corresponding simple random sample, the method in Verma, Scott and O'Muircheartaigh (1980) is used.

As can be seen in the table, in the case of the HBS. the values of *deft* range between 1.4 and 2.1 for income items, while between 1.5 and 2.6 for various groups of expenditures. The *deft* statistics for the LFS are smaller, in some cases values less than 1 were also obtained

Table 5. Design Effect of Some HBS and LFS Estimates from the Years 2000-2001

Denomination	Estimate	Standard Deviation		Deft	CV
		Jackknife	SRS		%
Average Household Incomes, in 1000 HUF					
Income from Work	949.532	22.180	11.081	2.00	2.34
Social Income	399.212	5.326	3.210	1.66	1.33
Other Income	23.590	2.091	1.502	1.39	8.86
Gross Income, Total	1372.334	20.900	10.111	2.07	1.52
Net Available Income	1126.662	14.972	7.350	2.04	1.33
Average Household Expenditures, in 1000 HUF					
Food	303.891	4.300	1.745	2.46	1.42
Beverages, Tobacco	57.109	1.021	0.614	1.66	1.79
Clothing	60.347	1.626	0.742	2.19	2.69
Maintenance of Dwellings	194.357	2.660	1.011	2.63	1.37
Furniture, Household Equipment	60.146	1.326	0.868	1.53	2.20
Health Personal Care	61.803	1.252	0.698	1.79	2.03
Transport, Communication	169.256	4.959	2.959	1.68	2.93
Culture, Recreation, Entertainment	69.571	2.564	1.195	2.15	3.69
Other Personal Expenditures	40.082	1.823	1.016	1.79	4.55
Investment on Dwelling	46.336	4.022	3.053	1.32	8.68
Personal Expenditures, Total	1062.899	15.210	7.590	2.00	1.43
LFS 1st Quarter 2001					
Unemployment Rate (%)	6.0	0.35	0.26	1.36	5.8
Unemployment Rate of Males (%)	4.1	0.30	0.30	1.02	7.2
Participation Rate of Aged	51.7	1.53	1.63	0.94	2.9

Denomination	Estimate	Standard Deviation		Deft	CV
		Jackknife	SRS		%
20-24 (%)					
Participation Rate of Females (%)	45.7	0.64	0.81	0.79	1.4
LFS 4th Quarter 2001					
Unemployment Rate (%)	5.6	0.36	0.25	1.41	6.4
Unemployment Rate of Males (%)	3.8	0.29	0.28	1.03	7.7
Participation Rate of Aged 20-24 (%)	49.8	1.63	1.68	0.97	3.3
Participation Rate of Females (%)	45.3	0.67	0.81	0.82	1.5

6. Household Budget Survey

Since 1976 the HBS is based on a random subsample of the USHS sample. The latter as sampling frame has always consisted of EDs and was updated after every decennial census. Between 1976 and 1982 the HBS was a continuous survey, between 1983 and 1991, however, was carried out only biennially. Since 1993 it is again a continuous survey.

An important aspect of the HBS sample that biennally the interviewers call every household in their EDs and collect some demographic and economic activity data about them (e.g. the size of household, age, educational level and economic activity of the head). These data are used only for *substitution* purposes and not in the course of the primary selection of dwellings from the ED. Namely, because of the rather high nonresponse rate one substitution (in Budapest and in the largest cities two) is allowed. The substitution household is selected from the same stratum of households the originally selected one belonged to and from the EDs belonging to the same interviewer. The address of the substitution household is given to the interviewer only after his/her reporting failure at the originally selected address. The yearly sample size is distributed evenly over the months. A household consenting to participate in the survey is asked to report its incomes and expenditures daily over a month. During this period interviewers collect certain additional data about the household (age and occupational structure of the household, type, size and equipment of the dwelling, stock of consumer durables, etc.). In addition, at the beginning of the next year the interviewer calls again the household to ask them -third of the households is *rotated* in such a way that the initial sample about some expenditures of high value during the whole year and certain types of yearly incomes.

Every year one size which is six times the number of EDs remains unchanged. Because nonresponding households may be substituted from all EDs belonging to the same interviewer, the actual number of households cooperating in the survey can be less or more than the initial six households. Thus the rate of

rotation in a given ED can be lower or higher than one third. However, a household participating in the survey through three consecutive years is rotated out after this period in any case.

In 2000 the HBS sample covered nearly 1980 EDs from 262 localities and the number of initially selected households was 11862. The structure of the sample by regions and size categories of localities.

As interviewers often encounter refusal or other type of nonresponse at the substitute addresses, too, the size of the realized sample is regularly smaller than the planned sample size. In 2000 e.g. instead of the 1977x6=11862 households only 10191 co-operated with the survey, but to achieve this result the interviewers had to call as many as 17243 addresses. *Nonresponse rates* increased from 1993 reaching 43.3% in 1996, then decreased to some extent. In 2000 the total nonresponse rate was 39 per cent and within this rate, refusals amounted to nearly 27 percent, the rest consisted of vacancies, not-at-homes, ceased addresses and other cases. The problem of nonresponse and to ensure the planned sample size is particularly difficult in the capital and in some large cities. Although co-operating households get a certain amount of money as *incentive* for supplying their data, unfortunately the amount is not large enough for many households to motivate them to co-operate with the survey. It must be noted, on the other hand, that a change in the remuneration system of the interviewers in the sense to inspire them to make more efforts in persuading households to co-operate in the survey considerably contributed to the fall in the rate of refusals from 34.4 percent in 1996 to 26.9 percent in 2000. Now, in the frame of a project for improving the reliability of the HBS data experiments are carried out to find out whether the reduction of the period of recording the expenditures from one month to, say, two weeks would result in some further decrease in the refusal rate.

The results of the survey are published yearly in a bilingual form (Hungarian and English) with a short analysis of the data. The next publication will appear in early 2002 by the title „Household Budget Survey, 2000 Annual Report" CSO Budapest, 2002. The publication is available also on a CD ROM.

It can be concluded that the HBS is one of the most important household surveys of the CSO. Its data are used not only to calculate weights for the consumer price index (CPI), but also to estimate the consumption of households within the frame of national accounts for producing the GDP values quarterly and yearly. In addition, its data are of vital importance for quite a number of institutions, researchers, universities to carry out research in fields like living conditions of various social strata, expenditure patterns of various types of households and changes in them, consumers' demand for different types of commodities, etc. For these purposes not only the official publications of the CSO are used , but the anonymized data file of the survey on a CD ROM can be purchased, too. It must be mentioned, furthermore, that in the course of harmonisation with the EU, from 2001 expenditures are grouped according to the COICOP system.

7. Labour Force Survey

LFS is a new household survey introduced by the HCSO in 1992. Its sample was selected in 1991 on the basis of the 1990 Census data. The initial sample in 1992 contained 9960 EDs in 670 localities and three addresses from each ED in a quarter of a year. This means that in a quarter 9960x3=29200 addresses were visited by the interviewers, i.e. on average almost 10 thousand households monthly. It must be mentioned that in the first few years of the LFS, randomly selected substitute addresses were given to the interviewers in case of refusal or other causes of nonresponse. The number of substitute addresses was equal to the number of primary ones, i.e. three addresses in a quarter. A few years later the system of substitution was abandoned and taking into account the experience concerning the nonresponse rates in different types of localities, the initial sample size was disproportionately increased.

Biennially the interviewers check the list of addresses in their EDs deleting those addresses which ceased to exist or are no dwellings any more and completing the list with newly built dwellings.

In the second half of the 1990s a demand for more detailed regional LFS data emerged and as reliability investigations showed that in many cases the sample size was not large enough to get reliable regional estimates, it was decided to *enlarge the sample* by about 30 percent from 1998. The enlargement was considerable in respect of the number of localities covered by the sample, too, and especially in the number of EDs, which increased by more than 50 percent. This latter resulted from restoring the original system of selecting three addresses from each ED quarterly. The structure of the sample by regions and size categories of localities in year 2000. More details about the enlarged sample can be found in Éltető (2000). From the mid 1990s no substitute addresses are used in the LFS.

Nonresponse rates in the LFS – especially rates of refusals – are much lower than those in the HBS. From the beginning till 1997 a slight increase in the nonresponse rates could be observed reaching a maximum of 14.3 percent, then total nonresponse rate tended to decline, it was 9.2 percent in 2000. Refusal rates also increased at first, reaching 7 percent in 1996 and 1997, then decreased, in 2000 only 3.2 percent of the selected households refused to co-operate with the interviewers.

Mainly because the sample of the LFS contain much more localities and especially EDs as PSUs and SSUs than that of the HBS, the *design effect* is considerably lower than in the case of the latter, see Table 5. In three of the four quarters LFS is in general supplemented by a *module* with special topics like the situation of working women, questions concerning mothers on child care leave, etc. One of the three modules, in general that for the second quarter of the year covers the topic recommended by Eurostat for that year. Both the basic LFS questionnaires and those for the Eurostat modules are fully harmonised with

Eurostat recommendations in the sense that all information required by Eurostat can be acquired from the questionnaires used in the LFS.

In 2002 a new sampling frame based on the data of the 2001 Population and Housing Census will be selected for the samples of the LFS. An important change in the sampling design is planned. As the HCSO is going to dispose of a yearly updated dwelling register utilising the data of the 2001 census and detailed yearly statistics on dwelling constructions and demolitions, there will be no need to select EDs as sampling units. Instead the whole dwelling base of the localities covered by the sample will serve as sampling frame for the LFS. However, possibly we are going to apply some kind of geographical stratification within the localities selected into the sample. That means that in the case of self-representing towns the sample of dwellings will be a stratified random sample, while in the case of non-self-representing localities a two-stage stratified sample. It is expected that this change in the sampling design will further decrease the value of the design effect.

Both quarterly and annual data of the LFS are published in bilingual bulletins.

Though the government receives information on the number of *registered* unemployed persons from other sources, too, the data of the LFS differ both in concept and in detail, especially concerning the family status, household composition and economic activity of other household members of the unemployed. Thus LFS yields valuable and indispensable information about the actual situation and changes in the labour market for both central and local governmental bodies as well as for researchers and other interested people. The official unemployment rate based on LFS data is one of the most important and monthly very much awaited indicator of the economy.

8. Other Household Surveys

8.1. The 2 Percent Microcensus in 1996

Based on the proposal of the president of the HCSO the Hungarian Parliament passed an Act in 1995 on the execution of a 2 percent Microcensus in April 1996. The Act made the response to the questions of the Microcensus compulsory. The Microcensus was based on two questionnaires: a dwelling and a personal one with questions similar to those generally used in censuses.

The sample of the Microcensus was a proportional sample, i.e. a general 2 percent sampling fraction was applied for both the population living in private households of the various geographic units or size categories of localities and that of institutional households. To select the sample three sampling frames were available:

- list of addresses of the 1990 Census
- list of dwellings built in years 1990-1995

- list of institutional households from an administrative survey carried out in September 1995.

Both organisational reasons and financial restrictions indicated that the sample should be selected in several stages. PSUs were generally the localities, while within them EDs served as SSUs and the addresses within them were the final sampling units. All localities with 15 thousand or more inhabitants were self-representing. It is a characteristic feature of the sampling procedure of the Microcensus that PSUs were selected with PPS, and in the non-self-representing localities an equal number of SSUs (EDs) were selected within a stratum from the selected PSUs also with PPS, while from the sampled EDs an equal number of addresses (generally 8, but 12 in case of localities with less than 1000 inhabitants) were selected with equal probability. This sampling procedure ensured that all addresses of the country enumerated in the 1990 Census had the same chance, i.e. 2 per cent of being selected into the sample of the Microcensus.

Before selecting the non-self-representing localities, they were stratified by the same size categories as in the case of the LFS, see Table 7. After all at national level the selection procedure from the list of addresses of the 1990 Census resulted in a sample consisting of 77,080 addresses from 9,382 EDs.

As to the sample from the second sampling frame we had statistics on the number of newly built dwellings in the years 1990-1994 and in the first half of year 1995 broken down by years, counties and cities and villages. The number of newly built dwellings in the first half of 1995 was doubled to estimate the number of newly built dwellings for the whole year 1995. From each group 2 percent of dwellings built in the years 1990-1995 was selected for the Microcensus, but only from localities already sampled. This part of the Microcensus sample contained 3,288 addresses of dwellings built after January 1, 1990.

Institutions were stratified into 5 size categories according to the expected number of persons living there at the Microcensus moment. Different sampling fractions were applied in the various size categories of institutions both relating to institutions and within them relating to persons in such a way that from each size category exactly 2 percent of persons should be selected. The largest 35 institutions were self-representing. Besides, institutions were stratified also by types to ensure that all of the eight types be represented in the sample properly. Altogether 310 institutions and from them 4,182 persons were selected into the sample of the Microcensus.

8.2. Surveys on Three 25 percent Subsamples of the Microcensus sample

The sample of the Microcensus was divided into four subsamples of equal size, i.e. each containing 0.5 percent of private households. Institutional households were not covered by the subsamples. The subsamples were formed at the level of EDs and addresses in such a way that from every second ED one half of the sample addresses belonged to a subsample.

On three of the subsamples three *module surveys were* carried out simultaneously with the Microcensus. Unlike the Microcensus itself these household surveys were naturally optional, i.e. responding was not compulsory. Still higher response rates were experienced than if the surveys had been executed separately.

Perhaps the most important one out of these three surveys was an *income survey*. Previously the last successful income survey of the HCSO was still before the change in the political system in 1989, in 1988, inquiring 1987 incomes, under quite different circumstances (e.g. the general income tax system was introduced in 1988 only). Experiences in the HBS as well as in other surveys showed that in the years of transition people had become especially sensitive to questions on incomes resulting in high nonresponse rates and considerable underreporting of incomes. Although nonresponse was rather frequent in the income survey connected with the Microcensus, the total and partial nonresponse rate was about 18 percent, quite a number of important information about nonrespondent households and persons (e.g. composition of the household, age, educational attainment, economic activity, occupation, industry of its members, size, equipment and quality of the dwelling etc.) were known from the Microcensus, which made it possible to apply well-founded imputations. As a matter of fact, for all nonrespondent earners it was possible to find a person (or several ones) with very similar characteristics among the respondents. Moreover, in case of certain income sources imputations were applied also on the basis of aggregate data obtained from tax authorities. For more details see Éltető-Havasi (2000). As a result the data of this income survey reflected reliably enough the income distribution, income inequality in Hungary in the mid 1990s.

On the second subsample a wealth survey was carried out. Interestingly enough the response rate was considerably higher than in the case of the income survey, perhaps because no question was put on actual incomes, but rather on the possession of consumer durables of high value, ownership of summer resort, secondary dwelling, land, art treasures etc.

The topic of third survey carried out simultaneously with the Microcensus was tourism at home and abroad. The survey was co-sponsored by several tourist offices. As the topic and the questions were rather neutral, the response rate was high. It is planned to carry out a similar survey on tourism in year 2002.

7.3. The Housing Survey in 1999

In May and June of 1999 a Housing Survey was carried out on somewhat more than a half of the third subsample of the Microcensus. One of the reasons to use a subsample of the Microcensus for the survey was the feasibility of making use of the data of the Microcensus on the dwelling, or to be more exact these data had to be inquired only in case of changes since the Microcensus. The sample was restricted to localities covered by the LFS sample in order to avoid the necessity of recruiting new, inexperienced interviewers. The sample was proportionately

allocated to size categories of localities as well as to the counties. The sample covered 2613 EDs in 357 localities and from these altogether 10761 addresses. However, because considerable nonresponse could be expected, a supplementary sample of EDs was also selected amounting to nearly 40 percent of the number of EDs in the basic sample to ensure the originally determined sample size. The supplementary EDs belonged also to the third or occasionally the fourth subsample of the Microcensus.

7.4. Time Use Survey in 1999-2000

Originally the Time Use Survey was planned in years 1996-1997, but for lack of the necessary financial funds it had to be postponed to years 1999-2000. The survey covered a whole year beginning in September 1999 and ending in August 2000. The sample of addresses was a part of the fourth subsample of the Microcensus. This had the advantage that for the selection of the person to be asked from each sampled dwelling to participate in the survey the age-sex composition of people living in the sample dwellings was available in advance and thus a computer algorithm could be elaborated to carry out the selection of persons in such a way that the age-sex structure of the sample could correspond exactly to that of the whole population. Each person co-operating in the survey was asked to report his/her activities by quarter of an hour on the designated day four times during the year – once in the same month (first, second or third) of the quarter. Besides listing the activities in detail there was a supplementary questionnaire containing the personal data of the sampled person and certain questions on his/her way of life.

Like the fourth subsample itself the sample was proportionately allocated to the various geographic units and size categories of localities. In this household survey, too, the sample was restricted to localities covered by either the HBS or the LFS sample in order to have experienced interviewers for the survey. The primary sample contained about 10 thousand addresses which was complemented further by one thousand ones for substitution in case of nonresponse. The sample covered 361 localities.

Nonresponse rate was highest in the first quarter, it was 31 percent, and out of it moving from the selected address to another one was the main cause of nonresponse, it amounted to 35 percent of all nonresponses. However, the age-sex structure of the originally selected sample and the actual one after substitutions were very similar, they did not differ in none of the groups by more then 2 percent.

9. Concluding remarks

In this paper a the past and present of the household surveys of the Hungarian Central Statistical Office (HCSO) are presented, with specific emphasis on two continuous surveys, namely, the Household Budget Survey

(HBS) and the Labour Force Survey (LFS). The HCSO has sound traditions and experience in conducting household surveys. This proved helpful at the beginning of the political and economic changes in the early nineties when a number of unusual phenomena such as e.g. unemployment manifested themselves and created thereby unexpected challenges for the Hungarian official statistics. To overcome the difficulties coming from deteriorating willingness of people to co-operate with the interviewers of the HCSO was a hard job, but the worst of the hardships is over. At the beginning of the 21st century, the HCSO has good international links, among other things to the Eurostat and Statistics Canada, and this facilitates for the HCSO staff to keep pace with recent developments in official statistics. Owing to these conditions, the standards of the Hungarian household surveys are proper for a country, which will join the EU.

REFERENCES

CHAMBERS, R.L., and SKINNER, C.J. *(1999). Intelligent calibration. Bulletin of the International Statistical Institute, Proceedings, LVIII, Book 2, 321-324.*

Council Regulation (EC) No. 577/98 of 9 March 1998 on the organisation of labour force sample survey in the Community. Official Journal of the European Communities, 14/3/98, I.77/3-I.77/7.

DARROCH, J.N., and RATCLIFF, D. (1972). Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics*, 43, 1470-1480.

DEVILLE, J.-C., and SÄRNDAL, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association* ,87, 376-382.

ÉLTETŐ, Ö., MIHÁLYFFY, L. (1997) Stability of Composite Estimators: Experiments with Hungarian LFS Data. *Hungarian Statistical Review,* Vol. 76. Special English Issue.

MIHALYFFY, L. (1998) A time series approach to adjusting data in longitudinal surveys. Paper presented at the Joint Conference IASS/IAOS, Aguascalientes, Mexico 1-4 September, 1998.

ÉLTETŐ, Ö., HAVASI, É. (1999) Income Inequality and Poverty in Hungary in the Mid 1990s. *Hungarian Statistical Review*, Vol.77. Special English Issue.

ÉLTETŐ, Ö. (2000) Enlargement of the Sample of the Hungarian LFS To Get Reliable Small Area Estimates for Labour Market Indicators. *Statistics in Transition*, Vol. 4, Number 4, March 2000.

VERMA, Vijay – Scott, Christopher – O'Muircheartaigh, Colm: Sample Design and Sampling Errors for the World Fertility Survey. *Journal of the Royal Statistical Society*, Ser. A (1980), Vol. 143, 431-473.

STATISTICS IN TRANSITION, June 2002
Vol. 5, No. 4, pp. 541–554

THE LABOUR FORCE SURVEY IN THE CZECH REPUBLIC

Martin Kopecky, Ivo Makalous and Zuzana Pavlickova[1]

ABSTRACT

The paper presents some aspects of designing and implementation of the Labour Force Survey (LFS) in the Czech Republic. The survey, started in 1992 as a continuous quarterly survey. The sample was selected in two stages stratified by regions. The primary sampling units were census enumeration areas, selected with probabilities proportional to the numbers of dwellings. At the second stage the same number of dwellings was selected (the number was changed in 2000). Weighting techniques are also described. Some examples of sampling errors are presented. The non-response rates, cost components of the survey, and other organisational aspects are also reported.

1. Introduction

The situation on the labour market before the transformation of Czech economy was monitored by business reporting only. However, this method was effective under full employment. With progressing transformation, the administrative sources of issued permits for enterprising and licences became available. On the other hand the net of labour offices started and the numbers of registered unemployed persons became available. From 1991 the labour force survey was prepared, it filled out the existing data and ensured the international comparability of Czech labour data.

The Labour Force Survey (LFS) in the Czech Republic has been conducted on a regular basis since December 1992. Based on the recommendations given by international organisations and following consultations with our colleagues from the OPCS (United Kingdom), LFS is a quarterly continuous survey from its beginning. Each selected dwelling is measured for five quarters. The own net covers 113 interviewers. Data processing is centralised into the unit that ensures all survey-related activities. The results are disseminated periodically every

[1] Czech Statistical Office, Prague, The Czech Republic.

quarter. The time series of basic labour market indicators are published irregularly.

The survey covers the population aged 15 years and above, and other persons are asked the demographic and technical questions to obtain complete household structure.

2. Sample Design

The LFS concentrates on households living in dwellings that have been selected at random. It addresses all persons living in the dwellings continuously for at least three months, disregarding the type of their stay there (permanent, temporary or non-registered). The exception is made for the temporary members of the armed forces who are surveyed, for practical reasons, at the dwelling where they lived before they left for the army. For persons under 15 years old, we ask only for the basic data: relationship to the head of household, age, sex, nationality and ethnicity. The survey does not cover persons living in any collective accommodation for a long period of time. That is why the data on certain population groups (foreign nationals living and working in the CR in particular) are rather scarce. According to Eurostat, such data can be retrieved from the administrative sources or acquired via a separate survey taken outside the LFS.

In the Czech Republic the Register of Census Areas is used as the sample frame. The data about houses and dwellings are permanently being updated based on the administrative sources. The average size of a census area is about 80 dwellings. The census areas with 15 dwellings and less (which represent 1.2 % of the total number of the dwellings in the CR) are not considered for sampling. It eliminates the cases, when in the selected census area there are not the necessary 6 dwellings for sampling on the second stage due to insufficient updating of the Register of Census Areas. The limit of 15 dwellings was set on the base of experience with updating the Register of Census Area and on the share of these census areas in the total number of dwellings. It means that by decreasing this limit the risk of inapplicable census areas would rise and by increasing this limit too many dwellings would be excluded from the sample frame.

After the execution of the survey the Register of Census Areas is updated based on this survey and the surveyed dwellings are not to be included into the sample again for 5 years.

The criteria for including the census area repeatedly into the sample on the first stage exist as well; they differ only in the dependency on their size. The census areas up to 50 dwellings can be selected again after 12 quarters, the ones from 50 to 100 dwellings can be selected at most twice every 12 quarters and those above 100 dwellings at most thrice every 12 quarters. Together with these criteria it is ensured that no census area may be included into the sample before it's excluding. It ensures that every quarterly sample has the same quality as the sample without replacement.

The sample is selected as a regional in two stages. On the first stage the census areas are sampled in every district by the method of randomised systematic sampling. The probabilities of inclusion are proportional to its size (PPS). On the second stage of the sampling 6 dwellings are sampled in every selected census area by the method of simple random sampling.

The final probability for dwelling k in the census area j in the district i depends only on the number of the sampled dwellings and on the total number of dwellings in the district i. See the formula below.

$$\pi_{ijk} = \pi_{ij}^{I}\, \pi_{k}^{II} = l_i \frac{D_{ij}}{D_i} \frac{6}{D_{ij}} = \frac{l_i}{D_i} \frac{6}{D_i} = \frac{d_i}{D_i}$$

D denotes the total number of dwellings, d stands for the number of dwellings in the sample and l denotes the number of selected census areas.

For the LFS sample the five-wave rotation scheme is used. It means, that every quarter 20 per cent dwellings of the total sample size are newly included into the sample and 20 per cent dwellings that have already been surveyed for 5 quarters are excluded. For each quarter the independent sample is done, so the quarterly sample consists of five independent sub-samples. This is important for computing the sample errors of annual averages. Within a year eight of these independent sub-samples are surveyed and the annual sample is greater than the quarterly sample in 60 per cent.

2.1. Sample sizes

Sample size on the NUTS 2 level corresponds to the requirements of Eurostat mentioned in Council Regulation No 577/98. It means, that the relative standard error of an estimate representing 5% of population in working age does not exceed 8 per cent. Those NUTS 2 designed from NUTS 3 have greater sample than is needed due to reliability of estimates on the NUTS 3 level.

The Eurostat does not have any requirements for sample sizes on the NUTS 3 level, so the adjusted requirement for NUTS 2 was used; the relative standard error of an estimate representing 5% of population in working age does not exceed 10 per cent. The sample sizes for both statistical units NUTS 2 and NUTS 3 were estimated on the assumption that the total non-response rate is 30 per cent.

The sample sizes in particular districts were evaluated so that the sample in every NUTS 3 unit is well proportioned. Only 4 districts have overrated the sample due to estimate of employment necessary for registered unemployment rate on the district level. The relative sample size after deducting non-response of dwellings in particular NUTS 3 units was moving from 4‰ (4.3 thousand) in the capital city of Prague to 9‰ (4.8 thousand) in Plzeňský region in the year 2000. Currently it is from 4‰ (4.1 thousand) in the capital city of Prague to 9,4‰ (2.8 thousand) in Karlovarský region which is the smallest region. On the level of the republic the average sample size is slightly above 6‰ which represents about 25

thousand dwellings. In these dwellings more than 63 thousand people of all age groups were surveyed (almost 54 thousand respondents aged 15 or more).

3. Weighting and Estimating

For weighting we use post stratification by sex and age groups, and the used weight is in the following form

$$w_i^{(\nu)} = \frac{N_i^{(\nu)}}{n_i^{(\nu)}} \ ,$$

where N is the size of the basic population and n is its sample size and ν indexes 24 age groups (12 for women and 12 for men). The number of persons in the age groups and districts is projected from the final states of the two preceding years to the centres of the current quarters. The projection is based on the population trends of getting older and the trends of growth and fall of the population in the preceding years. In the past the simple method of linear extrapolation was used but the forecasts for several quarters forward did not have satisfactory reliability. The number of people in the sample is a random value (the sampling unit is not a person, but the dwelling – therefore even the number of households in the sample is a random value) and the final estimates are biased. However, its efficiency is greater than for Horwitz-Thompson's estimates. The weighted biased estimate is in the following form

$$\overset{\vee}{Y} = \sum_{i \in District} \sum_{\nu \in Age\ gr.} \sum_{j=1}^{l_i} \sum_{k=1}^{d_{ij}} y_{ijk}^{(\nu)} \ \frac{N_i^{(\nu)}}{n_i^{(\nu)}} =$$

$$= \sum_{i \in District} \sum_{\nu \in Age\ gr.} \frac{y_i^{(\nu)}}{n_i^{(\nu)}} N_i^{(\nu)}$$

It is evident that the used estimate is a ratio estimate. Characteristics $y_i^{(\nu)}$ and $n_i^{(\nu)}$ are random values and represent numbers of people with certain characteristic in the post stratified sub-sample of the district i and age group ν and the total number in this sub-sample. This estimate is better than those obtained by means of the inverse probabilities of inclusion decreased by non-response are used. Moreover, it compensates the discrepancies between different non-response rates in the individual age groups.

All types of estimates (both qualitative and quantitative) are obtained from this formula. The only dealing with non-response consists in this post stratification. Other methods for adjustment of the data due to non-sampling errors have not been used in LFS so far.

4. Estimation of sample errors

Provided that the stratification and post stratification are assumed, the calculation of standard error is too demanding and its effect is almost zero. There is a relatively high non-response rate that makes non-sampling errors inaccessible. Further, it is necessary to mind that mainly qualitative variables are surveyed (y takes value 1 with probability p and value 0 with probability 1-p), which are a little sensitive in terms of stratification. Therefore it is assumed for the simplification that there is no stratification and the sampling unit is a person. In the quarterly published by the Czech Statistical Office "Employment and Unemployment in the Czech Republic as Measured by the Labour Force Sample Survey" there are simplified formulas for computing 95% confidence intervals and the tables for its estimate.

4.1. The data and sample quality

For data collection an electronic version of the questionnaire was created. There are defined about 200 checks of the internal relationships between the questions inside the program. The majority of them have the nature of notification only. It depends on the interviewer whether he/she confirms the notification or corrects the question. The subsequent checks are carried out by regional co-ordinators who monitor the correctness of those confirmations and other relationships that are not checked at the level of the interviewer. After combining the data at the regional levels the co-ordinators have immediately at their disposal the database to make the checking cross-tables. They serve for checking of the data quality as well as checking the interviewer's work as well. Another questionnaire checks are performed at the national level. Besides the check of the internal relationships between the questions, the chi-squared goodness-of-fit test on population distribution in the sample is executed (according to demographic data for particular districts).

As mentioned above, the adjustment of data due to non-response is not carried out. The attention is rather paid to decreasing the non-response rate and improving the data quality at the moment of their collection. This is also the reason why the system of reserve dwellings was cancelled (the interviewer had to compensate for the loss due to non-response of 5 basic dwellings from the 3 reserve dwellings in the census area). Although this system was increasing the response rate its impact on the quality sample was negative. It also was increasing the ratio of the households willing to participate in the survey to the others. Now only 6 dwellings in every census area are sampled and they are selected by the method of simple random sampling.

Table 1. 95 % confidence intervals for the number of employed in the national economy, unemployed and unemployment rates (4th quarter of 2000)

	Estimate	95 % confidence interval	Estimate	95 % confidence interval	Esti-mate	95 % con. interval

	(thou-sands)	Abs. -/+ (thou-sands)	Rel. -/+ (%)	(thou-sands)	Abs. -/+ (thou-sands)	Rel. -/+ (%)	(%)	Abs. -/+ (%)
	Employment			Unemployment			Unempl. rate	
Czech Republic	4751.0	36.6	0.8	430.4	15.9	3.7	8.3	0.3
Regions:								
NUTS 2 Praha	611.9	16.0	2.6	26.8	5.2	19.5	4.2	0.8
NUTS 2 Středočeský	520.9	11.3	2.2	40.6	4.6	11.4	7.2	0.8
NUTS 3 Jihočeský	301.2	7.4	2.5	18.1	2.8	15.3	5.7	0.9
NUTS 3 Plzeňský	269.0	7.0	2.6	15.1	2.5	16.7	5.3	0.9
NUTS 3 Karlovarský	152.8	6.4	4.2	11.6	2.8	23.7	7.1	1.7
NUTS 3 Ústecký	357.7	10.8	3.0	61.0	6.2	10.1	14.6	1.5
NUTS 3 Liberecký	204.3	7.9	3.9	11.6	2.8	24.4	5.4	1.3
NUTS 3 Královéhradecký	261.5	8.3	3.2	18.0	3.2	18.0	6.4	1.2
NUTS 3 Pardubický	228.9	8.2	3.6	18.1	3.3	18.4	7.3	1.3
NUTS 3 Vysočina	241.8	8.0	3.3	15.7	3.0	19.3	6.1	1.2
NUTS 3 Jihomoravský	519.0	11.8	2.3	47.5	5.2	10.9	8.4	0.9
NUTS 3 Olomoucký	277.7	9.5	3.4	38.6	4.9	12.7	12.2	1.6
NUTS 3 Zlínský	270.2	8.1	3.0	20.7	3.2	15.6	7.1	1.1
NUTS 2 Moravskoslezský	534.1	13.3	2.5	87.1	7.3	8.4	14.0	1.2

Source: LFS publication – Employment and unemployment in the Czech Republic as Measured by the Labour Force Sample Survey in 4th Quarter 2000.

The number of prescribed visits is the other method of decreasing the non-response rate. When a dwelling cannot answer the questions because the respondents have not been reached, the attempt at the contact is repeated. However, according to the rules, the maximum number of attempts is three. It is not allowed to visit the dwelling three times in one day. The attempts must be spread out into more days and more daytime. Only after three attempts (or in case of the refusal earlier) the interviewer can consider the dwelling as not reached. Unfortunately, despite all this effort the non-response rate remains relatively high.

5. Non-Response

As mentioned above, in a case of unsuccessful attempt at the contact of household there were predefined substitutes – the system of reserve dwellings. In 2000 this system was cancelled and the calculation of non-response rates has been made by the standard method since 2001.

In table below, the data for total non-response rate are reported. They are not fully comparable in time due to shift in reference quarter and changes in sampling method.

Table 2. Non-response rate in the period 1996-2000

Period	Autumn 1993	Autumn 1994	Autumn 1995	Autumn 1996	Autumn 1997	4th Quarter 1998	4th Quarter 1999	4th Quarter 2000
Non-response rate (%)	16	16	18	20	19	21	22	24

Source: own calculations

According to the first half-year of 2001, it seems that share of households which refused the co-operation is approximately 44% and share of households which were not reached by interviewers is about 36%. Administrative or technical waste, i.e. empty dwellings, dwellings not used for living and non-existent dwellings, cause the rest 20%.

Non-response is different in particular waves of visit, in the first wave, where a figure is the highest, does not exceed 30%. For the repeated visits it is considerably lower.

The non-response rate is influenced mainly by the willingness of household to participate. Technical and organisational conditions also play the role in, as mentioned in the chapter Survey Costs.

6. Experiences with Non-Sampling Errors

To verify the reliability of the sample survey, CZSO regularly analyse and compare LFSS figures and indicators with those based on the registers of labour offices.

On one hand, the differences of both figures due to distinct methodology are watched in detail (see the table below).

On the other hand the number of registered job applicants measured by the LFS and official number from labour offices is compared for data quality check. In the period in question the number of job applicants actually registered by labour offices reached 448.4 thousand, which differed by 32.1 thousand from that of the sample survey. This discrepancy has been appearing for a long time and we can assume that it is a non-sampling error (i.e. an error not caused by a statistical deviation between the characteristics of the sample and the basic sample). This is likely due to the fact that persons for whom it is difficult to find a job are hesitant to report it in the sample survey. In some cases it may also be caused by the fact that the respondent does not have to inform about being registered at an employment office, after conveying that he/she worked at least one hour in the reference week. Consequently, data on totals of the unemployed or registered job applicants may be underestimated in the sample survey.

Table 3. Breakdown of registered job applicants and the unemployed as identified by the LFSS (4th quarter 2000)

Category	Number (thousands)
a) Registered job applicants not available for work within 14 days	16.1
b) Registered job applicants not seeking a job during the last 4 weeks	19.2
c) Registered job applicants who worked at least 1 hour in reference week	6.1
d) Registered job applicants meeting ILO conditions for inclusion among the unemployed	374.9
All job applicants registered by labour offices (a+b+c+d)	416.3
e) The unemployed not registered, but meeting ILO conditions	55.6
The unemployed by ILO in total (d+e)	430.4

Source: LFS publication – Employment and unemployment in the Czech Republic as Measured by the Labour Force Sample Survey in 4th Quarter 2000

7. Survey Costs

In this section the used approaches for optimisation of dependency between quality and survey costs are brought to all parts. Finally the review of survey costs is stated.

The survey costs are very important for data quality, reliability and variability. Influence on money spent can be exerted by the following factors:

- the way of organising the interviewers' network,
- the used technology of data collection,
- the use of available technical means of data collection.

8. Organising the interviewers network

In principle there are two basic models of organising the interviewers network. The interviewers can either be contracted or professional. These two ways can be combined or different organisation may be used (e.g. an interviewer can do more jobs). Each of these two ways has its pros and cons and concrete conditions of the survey are decisive.

Contracted interviewers will be employed where there is not enough work to use interviewers capacity. This organisation is more economical in terms of cost of working force, but it must be complemented with a well elaborated and rather costly system of training given to a frequently changing team of interviewers. New interviewers have always to be initiated into the whole range of work (methodology, computers, software, nomenclatures, as well as the way of dealing with people, the importance of the survey, and many related issues). This type is particularly appropriate for non-continuous surveys (e.g. on an annual basis).

When the survey is non-continuous, but at a shorter periodicity (e.g. one reference week in a month), so that the interviewer has no other work to do for a certain period, we can take a compromising approach and fill in interviewers´ capacity with other activity and make him a professional. In any way, one needs an adequately higher number of interviewers for a non-continuous survey when it is necessary to fill in the sample in short time (e.g. within a week). This makes the survey more expensive. However, it may also result in quality lost for a higher number of people handling the survey and produce errors due to individual interpretation.

Professional interviewers are more appropriate when it is possible to fully use their capacity. This type of interviewers was being chosen in Czechoslovakia, where LFS was established as a continuous survey, i.e. conducted without interruption during the whole reference period, or during the whole year. That is why both Czech and Slovak Republics also now use a network of professional interviewers. Such a network has many advantages, above all the ones mentioned below.

- Constant monthly cost of labour. The cost of labour in case of contracted interviews mostly depends on the number of dwellings measured. For these reasons, extensive paperwork (admin) must be done. If the interviewers are professionals, only limited paperwork is required to ensure the technology of the survey.
- Basic training of all interviewers takes place at the same time. That is why training is not so expensive. Basic training has to contain e.g. survey methodology, proper acquaintance with the questionnaire including the meaning of individual concepts, work with nomenclatures, working with computer or notebook, software and work with sets, necessary paperwork, principles of effective contacts with people and many other skills that the interviewer must properly know how to use in the fieldwork. It is done by regional co-ordinators (14) with support of headquarters of the Czech Statistical Office.

The quality of the interviewer network (and subsequently the quality of data) can be gradually improved.

As each survey is developed (changes in methodology, new versions of software, new techniques and other changes are coming), it is always better to work with people who already know the ropes and understand the sense of changes introduced. Moreover, professional interviewers somehow perceive their share in responsibility for good results of the survey and a certain feeling of some importance in the team, which brings higher data quality and lower turnover of labour.

9. The technology used for data collection

The used technology of data collection can very significantly influence the costs of the survey. At the same time, it has to make account of the methodology of the survey and of the samples in general. While preparing our LFS, we tried to choose the least costly techniques, which met the required criteria. As it is impossible to mention all the rules, here are the most important ones:

- **Limited attempts at a contact.** When a dwelling cannot answer the questions because the respondents have not been caught, the attempt at a contract is repeated. However, according to the rules, the maximum number of attempts is three. When all the attempts fail, a dwelling is marked as not-at-home. This solution is necessary for two reasons: reducing the survey costs and time reasons, as the survey must continue to comply with plan.
- **Selecting the right member of the household.** When each member of the household is able to answer the question for him/herself, the results are most reliable. However, this principle cannot always be stuck to, as not all the household members are always available or ready to answer. For this reason, we apply a method where one household member is responsible for the household (when it is possible to do interviewers with individual members during one visit). This method enables us to minimise costs, i.e. particularly travel expenses. This household member must be the best informed one about all the others to provide reliable information.
- **Copying out answers.** When the respondent has not been caught even at the third attempt at a contact on repeated visit, the interviewer is allowed to copy out the previous information provided. However, this is allowed only once; when this is the case also in the following quarter, the dwelling is left out from the survey. This procedure was consulted with British experts, who are using it too. On our experience, copying out is applied particularly in holiday months and its share does not exceed 3 % of all interviewers.
- **Processing the survey results.** Another possibility of reducing costs is in the appropriate location of processing. A separate unit was established at the Czech Statistical Office already at the beginning of LFS that is the charge of methodology, organisation, analysis and publishing activities, and results processing. Apart from this, this unit prepares complex software, which has not be contracted out and paid for. Another advantages of this approach are its flexibility; relatively fast and reliable are not only the publishing of regular quarterly publications, but also the compilation of non-standard outputs. For our LFS the BLAISE, a system developed by the Dutch CBS, is being used.
- **The use of available technical means of data collection.** Interviewers who are our employees can make better use of available technical means to streamline their work (company cars, computers, data network, telephones, faxes, etc.). Technical means available for LFS in the Czech Republic are as follows:

Notebooks. All interviewers and regional co-ordinators are equipped with portable computers. Software for LFS has been prepared for both working with

paper questionnaire (CADI) and with a notebook (CAPI). When the interviewer does not use a notebook in the household, he/she puts the data into the computer subsequently on his/her own, so no other staff specialised in data acquisition is necessary. Another good point is that the interviewer remembers the content of the interview and is thus able to correct any inaccurate answers at data acquisition either immediately or on the basis of additional contacts with the household. Since 2002 the electronic questionnaire has been used exclusively.

The notebooks were replaced during 2000-2001. Therefore, today, it is possible to use 32-bit SW applications that are fully compatible with operating system Microsoft Windows NT, 98, 2000, … However, it is very difficult to find an appropriate type of notebook on the market, as there is the principle contradiction between the required low weight and high battery capacity. On top of this, our available recourses are limited.

Though, the investments on technical devices behove. In terms of notebooks, their economic return is up to 30 months. The reckoned safe for 1 dwelling when using notebook is around 12-14 minutes. The purchase price of adequate notebook moves around CZK 70.000. Exchange rate for CZK (Czech koruna) is 32 per 1 EUR. Sample for 1 interviewer covers 100 dwellings a month. The hour wage costs of interviewer are about CZK120. It results in the fact that the investment is refunded after 24-29 months.

Telephones. Telephones help to streamline the survey too. Interviewers particularly for repeated interviews use them when this is agreed with the household. For the first contact a visit in person is required. However, these calls are no classic type of CATI, because it is not allowed by our technical conditions. Telephone calls are used above all in large towns and cities with a fairly dense phone network. Interviews by phone represent roughly 50 % on repeated visits. They enable us to save time and reduce travel expenses. We consider about establishing a call-centre with standard usage of CATI survey technique in the future.

Travel. Besides salaries, journeys to respondents constitute the largest portion of survey costs. Most often cars with drivers are used (accounting for about two thirds of travel expenses), but they are not available all the time and in all places. Most statistical units in the districts have one car available, which is shared for the purpose of other activities. There are also cases of two districts sharing one car. However, it is often the only way of getting the interviewer to respondents, doing interviews in the required number of dwellings in a relatively short time, as the public transport is being reduced. When possible, the means of public transport are used; this is the cheapest way (some 5 % of travel expenses). The best solution to this problem is the interviewer's own car (must be insured), or a car owned by the statistical agency driven by the interviewer (which is not available everywhere). Generally, the transport by cars is the most expensive. Travel by car is the most economical way for less accessible places as the hour costs for the interviewer's standing time when using the public transport are

higher than costs saved on car. City public transport is often used in large towns and cities. In those cases, the statistical agency makes a contribution to the purchase of a seasonal ticket by an interviewer.

Data transfer networks. Data transfer networks constitute an irreplaceable element in the LFS system. They allow the flexibility of mutual contracts between regional statistical bodies and headquarters of the Czech Statistical Office.

Table 4. Structure of cost in thousands CZK per year, perinterviewer

	Length of using in years	**Total per year**
Investment expenditure - total		2600
notebooks	5	2275
mobile phones	4	325
	No of employees	Amount in thousands CZK
Material expenditure - total		5551
Mobile phones	130	858
Transport by own car	40	1272
Transport by referent car	40	989
Transport with a driver	25	600
Mass transport	25	159
Daily allowance of interviewers	130	1392
Daily allowance of drivers	25	156
Others	130	125
Salaries - total		29880
Salaries of field employees	130	28080
Salaries of drivers (40%-time worker)	25	1800
Total		38031

Source: own calculations

10. Unemployment and General Unemployment Rate

The number of the unemployed derived from the LFSS results stood at 430.4 thousand in the 4th quarter 2000, including 233.9 thousand females (54.3). The numbers of unemployed women are higher than those of men in almost all age groups of working age. This is especially true for those aged 25 years and more. The number of unemployed handicapped persons was estimated at 37.3 thousand (8.7% of total unemployment).

The number of unemployed with primary education amounted to 116.8 thousand (i.e. 27.1% of total unemployment), and the number of unemployed dropouts from secondary schools without the General Certificate of Secondary Education was also high (42.9%). Taking into account educational attainments,

the structure of unemployed men is somewhat different from that observed for unemployed women. While persons with primary education and trade schools dominate among unemployed men with 27.8% and 48.7%, respectively (26.6% and 38.0%, respectively, among unemployed women), also persons with secondary education with GCE are strongly represented among unemployed women (almost a third - 31.7% of unemployed women).

The number of persons unemployed for a long time (i.e., by ILO definition, persons searching actively for a job for more than one year) reached 213.0 thousand and is steadily growing (per year by 24.0 thousand). This group was strongly represented by persons with basic education and without education (third of cases).

General unemployment rate (ILO) revealed by the LFSS was 8.3%. The rate of registered unemployment based on registration by labour offices was a little higher in the same period - 8.6%.

11. Included Modules

The LFS net of interviewers is easy to use for further or complementary surveys in households. When using same sample simultaneously with standard LFS module and small range of additional questionnaires, the both material and wage survey costs are minimal. Thus, it seems that using the LFS net with all human and technical resources is the best solution. The standard module burdens every interviewer for the full-time, i. e. to ensure not exceeding the legal limit of150 hours a year for after hours work.. The fundamental restriction of this solution is time spent in the household. The interview usually takes more than 1 hour if the household is economically active, participates in education or training and has more than one member. It is unbearable to execute another survey in this household. Therefore, we seek to reduce the sample of additional survey to repeated visits when the time of interviewing the household is shorter. Czech Statistical Office has executed several additional surveys in this turn:
1994 – The survey of dwellings (form of ownership, costs for living)
1999 – The survey of dwellings (form of ownership, costs for living)
2000 – LFS ad hoc module on Transition from school to working life
2001 – LFS ad hoc module on Length and pattern of working time

12. Conclusion

In this paper we tried to summarise some approaches of Czech statistics to collecting the labour market data through the LFS. This survey is considered to be the basic source of information in this field and it provides irreplaceable data for users. During the transformation period, the essential task is harmonisation of the national survey with standards used in the EU. The harmonisation proceeded

gradually and covered the number of steps concerning methodology of indicators, introducing new questions, definitions of basic economic categories, sample size and sample distribution, organisation of the survey and net of interviewers, processing and disseminating data. The last and most demanding change was made in 2001. From 2002 the Czech Republic will provide fully harmonised LFS data within the full range of compulsory variables with adequate technical characteristics.

REFERENCES

CZECH STATISTICAL OFFICE (2001), Employment and unemployment in the Czech Republic as Measured by the Labour Force Sample Survey, Prague, (published a quarter after reference quarter).

COUNCIL REGULATION (EC) (1998), No 577/98 of 9 March 1998 *on the organisation of a Labour Force Sample Survey in the Community.*

STATISTICS IN TRANSITION, June 2002
Vol. 5, No. 4, pp. 555–589

THE HOUSEHOLD SAMPLE SURVEYS IN POLAND

Jan Kordos[1], Bronisław Lednicki[1] and Małgorzata Żyra[3]

ABSTRACT

The authors present the design and implementation of the household sample surveys carried out by the Central Statistical Office of Poland (GUS) in the last decade. i.e. during the transition period. The paper starts with a general description of household surveys carried out by the GUS before the transition period, followed by the design and implementation of the continues household surveys, i.e. the household budget surveys (HBS) and the labour force survey (LFS), and ad hoc surveys, such as living conditions surveys, health status surveys, demographic surveys, and other social surveys. Special attention is paid to sample designs, estimation methods, sampling errors, design effects, costs of the surveys, nonresponse rates, and dissemination of the results. Future improvement of household surveys after the Population Census 2002 is stressed.

Key words: Sample design, Household survey, Household budget survey, Labour force survey, Nonresponse rate, Design effect, Cost of the survey, Dissemination of the result.

1. Introduction

In order to provide sources of social data that are both more frequent and more comprehensive than censuses, the Central Statistical Office of Poland (GUS) has developed household sample survey programme. Here special attention is given to household surveys carried out in transition period, i.e. in years 1991-2000. However, different household sample surveys were conducted before the transition period, and general characteristics of those surveys are given below.

[1] Warsaw School of Economiss, and Central Statisticcal Office, Al. Niepodleglosci 208, 00-925 Warsaw, Poland. E-mail: j.kordos@stat.gov.pl
[1] Central Statistical Office, Al. Niepodległosci 208, 00-925 Warsaw, Poland. E-mail:B.Lednicki@stat.gov.pl
[3] Central Statistical Office, Al. Niepodległosci 208, 00-925 Warsaw, Poland. E-mail: M.Zyra@stat.gov.pl

Household surveys in Poland have a relatively long tradition and there is a comprehensive professional literature on this subject. The first household surveys were started even before Poland regained its independence in 1918 (Szulc, 1967). A considerable development of household surveys, especially in the field of population and living conditions took place in the inter-war period (GUS, 1930; Luszniewicz, 1959; Szulc, 1967; Wojtyniak, 1958). At that time household budget surveys and other questionnaire surveys on living conditions of the population were started.

After World War II the Central Statistical Office started to reconstruct and develop household surveys, but in the years 1950-1955 the surveys were discontinued, especially the household budget surveys. The surveys were gradually reintroduced after 1956 (GUS, 1970, 1976, 1986; Kordos, 1982, 1996; Luszniewicz, 1959, 1982) and the improvement of the methodology of those surveys and the extension of their scope took place.

In the eighties the so-called Integrated System of Household Surveys (ISHS) was gradually implemented. It was launched in 1982 and completed in 1992. The ISHS was understood as a set of permanent, recurrent and one-time surveys related to households, implemented according to a plan and to the users needs. It was based on the following (GUS, 1987; Kordos, 1985):

- Harmonised concepts, definitions and classifications,
- "Master sample" i.e. the set of sub-samples used in surveys,
- Territorial organisation ensuring an adequate quality of the results due to the adequate training of the in-field staff.

The most important component of the ISHS was the household budget survey based on a two-phase sampling, quarterly rotation of households within a year, one-third rotation of households in the three following years (this means that the two-thirds of households were included in the panel for four consecutive years) and a four-year cycle of the survey of sub-samples. This method of surveys was discontinued in 1992. At the same time, i.e. in the years 1983-1992 sub-samples selected for the household budget survey were used for over 30 social surveys by the introduction of the so-called modules depending on the users needs and the programme of recurrent surveys. At certain intervals the sample surveys of living conditions of the population were conducted, in which the size of the samples was about 100 thousand households (Kordos, 1974, 1985, 1987).

In the previous period, i.e. up to 1990 and also later many social surveys were conducted, in which sub-samples of households were used. Among them the following are especially worth mentioning: the social and occupational status of women (1983, 1987, 1991 and 1994); the situation of the elderly (1985, 1989); the situation of multi-children families (1985, 1990); the situation of young couples (1984, 1988, 1992); participation in culture (1985, 1990); participation in tourism (1986, 1992).

Among one-time surveys conducted in the previous period one could mention: survey on the youth (1987); family health status (1987); life paths

(1988); the quality of life (1985); phases of family development (1988); use of health care services (1989); durable goods at households (1990); nutritive needs of the population (1986).

The attempts to integrate surveys of households conducted in the eighties considerably facilitated the adjustment of household surveys to the European standards. This does not mean that all problems were solved. Further integration and improvement of the methodology of household surveys is needed.

2. Household surveys in transition period

Methodology and implementation of household sample surveys carried out in transition period, i.e. in 1991-2000, are described in more detailed way. The surveys were considerably extended and modified compared to the period before 1990. The household budget survey was and still is being improved, and in 1992 for the first time a new survey of labour force i.e. labour force survey (LFS) was introduced. Also other new sample surveys of households were launched such as: survey on the well being of households, health status of households and multi-aspect survey of the living conditions of the population.

We start with the presentation in a synthetic form of the household budget survey, which is one of the most difficult, and the labour force survey. Other one-time household surveys are also described generally. Next special attention will be paid to some methodological aspects common to all household surveys, such as sampling frame construction, a sample design, method of estimation, sampling error, costs of the survey, design effect, nonresponse, and dissemination of the results.

2.1. The Household Budget Survey

As already mentioned, the household budget surveys have nearly 45-year tradition (GUS, 1986, 1999; Kordos, 1982, 1996; Lednicki, 1982; Panek, 1990). A lot of attention was devoted to these surveys due to their special role in the analysis of the living conditions of the population. Various methods of surveys were experimented with and attempts were made to improve their organisation. At the beginning of the nineties the methodology of the survey was changed. In the new method of conducting the household budget survey introduced in 1992 the classification of incomes and expenditure as well as the classification of socio-economic types of survey was changed. A monthly rotation of households was applied, and for the second year of the survey 50 per cent of households are rotated and other 50 per cent participate in the survey in the same month as in the previous year. The same procedure was accepted in the third and forth year of the survey (to be more clear – we kept a panel of households during four consecutive years, i.e. we have a split panel). For the first time all types of individual households in Poland were included in the survey, which covered about 32 thousand of households. As in the method previously applied, the so-called

modular surveys on sub-samples of households participating in the household budget survey were conducted. The first cycle of such surveys conducted according to the new method was completed on the first sub-sample in 1995 (next in 1996). In 1996 the second cycle, which lasted till the year 2000, was started. In 2000 redesign of the HBS was done and some methodological aspects were changed.

The survey provides a comprehensive set of data used in various types of social analysis, and first of all, in major CSO publications (e.g. GUS, 2001a, 2001b). In 1997 again actions were taken up on the *integration of household surveys*[1]. Thus further improvement of the household budget survey and its integration with other household surveys is planned. This course of activities is consistent with the Eurostat recommendations (Eurostat, 1995, 1997).

Modular surveys. The new method of household budget survey also covers conducting modular surveys. In the years 1993-2000 the following surveys were conducted (year of the modular survey in parenthesis): social assistance (1993); nutritive needs (1993); housing conditions (1993); educational needs (1994); households in the market economy (1994) health status of the population and expenditure on health service (1994) durable goods in households (1995), needs of families and their satisfying (1995), participation in tourism and culture (1995); use of VCR; health care in households (1998, 2000); tourism and recreation (1998), personal and property insurance of households (1998), individual consumption of food (2000). In the modules introduced needs resulting from the adjustment with the European standards are also taken into account.

2.2. The Labour Force Survey

The survey on the economic activity of the population was implemented in Poland for the first time in May 1992 and was repeated on the quarterly basis till the 3rd quarter 1999. It was prepared according to the ILO recommendations and could be treated as a modern statistical survey (Szarkowski and Witkowski, 1994). In each quarter about 22 thousand households and persons aged 15 and above whom were members of those households were surveyed. The results of the survey are used in the broad scope for the evaluation of the situation on the labour market and the size of unemployment. Occasionally modules on selected social topics are included in the survey, which considerably extends the use of the results in social and economic analyses.

Modular surveys. Until recently, the following modular surveys, which considerably extended the scope of obtained information were conducted: socio-economic status of the unemployed (1993) rural labour market (1993), effectiveness of the labour market policy (1994 and 1996), professional career of

[1] See: Internal regulation No. 20 of the President of the Central Statistical Office of 30 October 1997 on the establishment of the Working Group for the improvement of the methodology and integration of household surveys.

the graduates (1994 and 1997), unregistered labour (1995 and 1998), the situation at the labour market and the living conditions of the disabled (1995, 1999), travels of the disabled (2000).

The survey results are processed and published quarterly. Redesign of the LFS took place in 1999 to adjust the survey to new administrative division of the country and improve its efficiency. A new sample design and estimation method is described in 3.2.2.

2.3. The 1995 Micro-census of Population and Housing

Several ad hoc sample household surveys were conducted in the last decade, the largest of which was the Micro-census 1995.

A micro-census, i.e. large-scale sample survey of population and housing provides data on the demographic structure and social and occupational status of the population, their education and information on households and families of various profiles.

In May 1995 a large-scale sample survey (micro-census) of the population and housing was conducted (Bracha, 1996; GUS, 1998a; Szwalek, 1996). This one was the third micro-census; the two pervious ones were conducted in the years 1974 and 1984. The census provides major data indispensable in social surveys, the long tradition of which does not require more extensive explanations. It should be added that the censuses provide an opportunity to capture data on the disabled, migrations and other additional topics in the field of social statistics.

The 1995 Micro-Census covered 5 per cent of population, i.e. nearly 600 thousand households. The sample design, sample allocation and estimation method, and standard error calculation are described in 3.2.3.

The last complete Census of Population and Housing was conducted in 1988, and the next one in May 2002.

2.4. Living conditions surveys

Since 1997 it was decided to conduct, besides the household budget survey, a multi-aspect survey of the living conditions of the population (GUS, 1998c). The survey was carefully prepared in co-operation with the INSEE experts and conducted on a large sample for the first time in mid-1997. The survey was repeated on a smaller scale each year using panel sub-samples, and on a larger scale every few years, thus it is worth presenting it in more detail.

The survey referred to both the material status of households (such as financial situation, of household assets, satisfaction of the major needs, use of social assistance) as well as non-material aspects influencing the level of living (such as the health status, occupational status, social integration, patterns of spending leisure time).

The objectives of the survey were the following:

- Obtaining information on the distribution of the living conditions of the population,

- Finding most vulnerable groups,
- Research on the reasons of the relatively worse situation of selected groups,
- Investigating relations between particular aspects of the living conditions,
- Research into the territorial distribution of the living conditions taking into account especially the situation of 5 selected voivodships.

A unit of the survey was a household and in the household a selected adult aged 18 and above.

The survey did not cover population living in collective and special households (such as in student's hostels, social welfare houses), households of foreigners and the homeless.

A sample of households was selected in a two-stage stratified sampling. Prior to the sampling the first-stage units consisting of census enumeration areas (CEA's) were stratified by voivodship and rural/urban areas. In each stratum the CEA's were selected with the probability proportional to the number of dwellings and then 6 dwellings were selected in the chosen CEA. In a case when a selected dwelling was inhabited by more than one household, only one of the households was selected. Next, in each household one adult aged 18 and above was selected. In total in the whole country 15 thousand dwellings were selected, but a half of the sample was located in 5 chosen voivodships, each consisting of 1.5 thousand dwellings. The remaining voivodships were treated as a separate group.

In total 12,524 households took part in the survey and responses to individual questionnaires were received from 12,371 persons. The response rate in case of households was 87%, and for adult persons - 86%.

In mid-1998 the survey was repeated on a smaller scale. The objective of the survey was not only obtaining the results for the year under observation, but also conducting comparisons with the preceding year. That is why it was decided to observe a sub-sample of the same households, i.e. about 2,250 dwellings, which took part in the survey in the previous year, which constitutes 50% of the surveyed sample in that year and the remaining 50% of the sub-sample was selected independently.

The sample for 1999 consisted of two sub-samples: the sub-sample selected in 1998 (panel) and a new sub-sample, the size of which was equal to the 1998 panel sub-sample. In this way in each year there were a panel sub-sample and a new sub-sample selected from the updated sampling frame. It means that a sub-sample selected in a given year as a new sample participated in the survey in the next year as a panel. This approach is a compromise, which allows achieving the objective of the survey: evaluation of parameters in the year under observation and the evaluation of changes as compared with the preceding year. In the survey about 4,000 household participated. A new large-scale living conditions survey was conducted in 2001, and size of sample was above 24,000 households, and 18,052 responded (nonresponse rate 25 per cent).

2.5. Population health status survey

The survey was conducted in April 1996. It covered 19.2 thousand households, in which 62,74 6 persons were surveyed: 47,924 adult persons aged 15 and above and 14,822 children up to 14 years old (Bracha, 1998). The response rate was 88.6% households in the dwellings selected in the sample. This was the first survey of the health status of the population in Poland conducted on such a large scale.

The major objective of the survey on the health status of the population was to obtain information on the health status of the population and health protection in relation to the demographic and social characteristics of persons, family situation, material and occupational status and place of residence. Due to the needs of the major sponsor of the survey (The State Fund for the Rehabilitation of the Disabled) the questions on the disabled were considerably extended.

The scope of topics covered three major groups of questions:

- broadly understood self-assessment of the health status,
- use of medical services,
- selected elements of lifestyle.

The health survey of the population was based on the WHO recommendations which allows to compare the results with other European countries, especially the EU Member States and the countries of the ECE region.

At the same time the results of the survey for 1996 create an information database on the health status of the population and selected aspects of health care. The data can be used as a comparative basis for the monitoring in the following years of the health effects achieved as a result of the National Health Programme. A few publications have been already prepared from this survey (GUS, 1997a, 1997b).

2.6. Time use survey

The Central Statistical Office conducted the recent time use survey in 1984. It was the third survey conducted in Poland at a larger scale. The previous surveys were conducted in 1969 and 1976 (Kordos, 1988). The next survey was planned for 1992, but due to financial reasons it was postponed several times.

The CSO prepared time use survey taking into account the EU standards. In June 1995 an international scientific conference was organised which was devoted to methodological issues of the time use surveys and analyses. The organisers of the conference, apart from the CSO, were Eurostat and the International Association for Time Use Research. At the conference proposals were presented concerning the methodology of the European Harmonised Time Use Surveys recommended by Eurostat. The results of the conference were published in a special issue of *Statistics in Transition* (vol. 2, No. 4, November 1995).

In 1996 the CSO conducted a small-scale time use survey on a random sample of 1000 households which covered persons aged 10 years and over. The

objective of the survey was, among others, the verification of the methodology proposed by Eurostat (GUS, 1998b). A large-scale time use survey was postponed

2.7. Panel surveys of households

The CSO conducted panel household budget surveys in 1982 -1992. However, those surveys were considerably different from the European Community Household Panel (ECHP) conducted by Eurostat (Kordos, 1992; Verma, et al., 1996). The ECHP was designed mainly to provide economic and social information at the individual level. It was also designed to obtain data for the population by an adequate change of the sample to obtain comparability between countries, due to the fact that standard sample designs and common operational methods are used to provide a general multi-dimensional picture by simultaneous observation of various variables for the same group of units. In addition the ECHP covered the basic part of information required for the construction of a relevant set of social indicators and for the publication of periodical social reports.

In 1994 the CSO jointly with the Warsaw School of Economics and the Ministry of Labour and Social Policy conducted pilot panel surveys, and next four rounds of the core survey in May and November 1995 and May and November 1996 (Panek, 1996). They referred to the living conditions of households, poverty and risk of poverty. Those surveys also differed from the ECHP. They collected information on households as a whole, while the ECHP in its prevailing part refers to members of the households.

The CSO has started methodological and organisational activities to harmonise its panel survey with the Eurostat requirements. For the 1994-2001 period, the ECHP survey was used to fulfil different political needs. But in view of the need to update the content according to the new political demands, and of the request for operational improvement, i.e. mainly the timeliness of the produced data, the replacement of the ECHP after 2002 was decided (Eurostat, 2001). Starting from 2003 a new survey is to be introduced for EU countries, i.*e. Survye on Income and Living Conditions* (EU-SILC). The cross-sectional and longitudinal micro-data sets will be updated on a yearly basis. Modules will be added to the cross-sectional component starting from 2004. The CSO of Poland is going to join the project in due time.

3. Common methodological aspects of household surveys

Some methodological aspects are common to nearly all household surveys in Poland. For this reason some aspects of sampling frame, sample designs, estimation methods, sampling errors, costs of the survey, design effect are discussed jointly.

3.1. Sampling frames

It is worth reminding that the sampling frame identifies the units from which a sample can be selected, either explicitly or implicitly, and procedures that account for all units of the survey population. Population censuses and documentation prepared for their conduction are the base for a sampling frame construction for household surveys in Poland. The census documentation can serve for creating primary sampling units (PSU's) using enumeration statistical districts (ESD's) or census enumeration areas (CEA's), usually adjusted to specific demands of a survey, and for secondary sampling units (SSU's) dwellings are usually used. Dwellings in ESD's or in CEA's are updated on annual basis and the updating covers: an increase of the dwelling stock due to the completion of new buildings, a decrease of the dwelling stock due to the demolition and changes in the boundaries of districts due to the changes in the administrative division of the country. For each district the sampling frame contains information on the addresses and estimated data regarding to the number of population and number of dwellings (GUS, 1998a) . We will characterise the ESD's and CEA's generally. At the end of 2000 there were 33,023 ESD's and 179,721 CEA's. Their sizes differ considerably: there were 898 CEA's without population, 1,141 CEA's with collective households and 178,580 CEA's with private dwellings. On the average, there are 5.4 CEA's per ESD. Distributions of ESD's and CEA's by number of dwellings are given in table 1 and table 2 respectively.

Table 1. Distribution of ESD's by number of dwellings at the end of 2000

ESD's	ESD's by number of dwellings						
	to 200	**200-400**	**400-600**	**600-800**	**800-1000**	**1000 and above**	***Total***
Number	4,774	15,789	10,056	2,244	158	2	**33,023**
Percentage	14.4	47.8	30.5	6.8	0.5	0.0	**100.0**

Source: own calculation from the CSO data.

Table 2. Distribution of CEA's by number of dwellings at the end of 2000

CEA's	CEA's by number of dwellings							
	to 20	**20-40**	**40-60**	**60-80**	**80-100**	**100-130**	**131 and above**	**Total**
Number	12,688	22,861	43,640	44,353	31,092	19,457	4,489	**178,580**
Percentage	7.1	12.8	24..5	24.8	17.4	10.9	2.5	**100.0**

Source: own calculation.

As one can see below, to prepare PSU's for sample selection of the HBS and LFS it was necessary to merge neighbouring ESD's or CEA's to get required sizes of PSU's. For example, for preparing PSU's for HBS from 33,023 ESD's (from urban areas PSU's had at least 250 dwellings, and for rural areas – 150 dwellings) 29,172 PSU's were constructed.

3.2. The household survey sample designs

The sample design specifies the way in which the elements on the sampling frame are to be selected. It is essential that every element on the frame (and hence every element in the population) has a positive probability of selection. The actual selection of the sample, which we refer to as a realisation of the sampling process, must follow the rules specified by the design. By a sample plan we mean a sample design, methods of estimation of parameters and standard errors.

3.2.1. Sample designs for the HBS

Different sample designs for HBS were applied in the last forty-five years (GUS, 1986, 1999; Kordos, 1996; Lednicki, 1982). Here we present the sample design of HBS which refers to the years 1996-2000 (GUS, 2001a). It assumed sampling of primary sampling units (PSU's) for a period of four years, and two sub samples were selected, out of which the first was selected for a given year and the second for the following year.

First sub sample was selected for the years 1996-1999 (and replaced the first sub sample used in the years 1992-1995) and the second sub sample was selected for the years 1997-2000 (it replaced the second sub sample for the years 1993-1996).

The sample was selected in a two-stage stratified sampling. The primary sampling units were ESD's or clusters of ESD's covering at least 250 dwellings.

In order to use this file as a sampling frame for a selection of a sample for the HBS, ESD's covering less than 250 dwellings were combined with the neighbouring districts i.e. those districts, which had adjacent numbers. The PSU's created in this way were stratified by voivodship and within each voivodship in two strata: urban and rural. In total 98 strata were created.

Sampling of the first stage units

When accepting sample design for the selection of a sample for the HBS an assumption was made that the selected sample should be more or less self-weighting. The PSU's were selected with the probability proportional to the estimated number of dwellings in a PSU. In the sample 1350 PSU's were selected, i.e. two sub samples of 675 PSU's each. The primary sampling units were selected separately in each stratum. The procedure of systematic sampling was applied after random ordering of the units (PSU's selected with PPS using the Hartley-Rao method). The sample selected in a given stratum was divided at

random into two sub samples, the first of which was used in the survey since 1 January 1996 and the second since 1 January 1997. As it was already mentioned, the first sub sample was used in the years 1996-1999 and the second in the years 1997-2000.

Sampling at the second stage - sampling of dwellings

When sampling units at the second stage the following organisational and methodological approach was applied:

a) the sample consists of two parts: permanent and replaceable, replaced each year,
b) in the survey the pattern of monthly rotation was applied, i.e. in each month a new sub sample was used,
c) in a given PSU each month 2 dwellings were selected in which all households were surveyed,
d) sampling of dwellings for the permanent sample was done at the beginning, i.e. for 4 years,
e) sampling of dwellings for the replaceable sample was done twice;

   – for the first sub sample:
   for the years 1996-1997 in 1995
   for the years 1998-1999 in 1997

   – for the second sub sample:
   for the years 1997-1998 in 1996
   for the years 1999-2000 in 1998;

f) due to the application of the sequential sampling in case of non-response a reserve sample of dwellings was selected which were arranged at random.

Due to the application of the above mentioned approach, the sampling of dwellings was done in the following way:

- dwellings in a given PSU were arranged at random,
- the permanent part of the sample covered the first 12 dwellings,
- additional dwellings were selected to the replaceable part of the sample, i.e. 12 dwellings twice and after 2 years again 12 dwellings twice,
- the consecutive *n* dwellings arranged at random constituted a subsample of dwellings for sequential sampling, where n is less or equal 150.

Redesign of the sample in 2000

In 2000 two sub samples of HBS were used:

a) first (675 PSU's) for years 1997-2000,
b) second selected in 1999 (1750 PSU's) for years 2000-2001.

Scheme of selection of both sub-samples was similar, but differences, except size, were connected with:

- size of PSU,
- allocation of PSU's in voivodships,
- details of stratification.

The sample design of the first sub-sample was described above. It was done for the former territorial organisation of the country (49 voivodships and division by urban and rural areas, i.e. 98 strata).

Starting from the 1st January 1999 a new administrative division of the country has been introduced. Poland was divided into 16 voivodships and 373 poviats (counties).

For selection of the second sub-sample, it was assumed that PSU's in urban areas should have at least 250 dwellings, and in rural areas – 150. These differences are connected with smaller nonresponse rates in rural areas. Allocation of 1750 PSU's in 16 voivodships was done in such a way, as to get approximately the same precision for each voivodship (Lednicki, Wesolowski, 1994). Before selection PSU's were stratified separately in each voivodship, according to the class of size of localities. Larger towns usually were accepted as separate strata.

Sampling frame for the secondary sampling units were dwellings in selected PSU's, prepared especially for this purpose. Selection of SSU's in the first sub-sample has been described above. For the second sub-sample in each PSU 12 dwellings were selected, i.e. one for each month in the year for the survey in years 2000-2001. Additionally, in each PSU 150 dwellings were selected independently for reserve sub sample, to be used in the case of nonresponse. The dwellings in the reserve sample were ordered randomly to be used in the survey.

The year 2000 was exceptional. First, a purpose of the redesign of the survey was to get accepted precision of the results not only for the country as a whole but also for voivodships. For this reason second sub sample was increased from 675 PSU's to 1750 with sample allocation to get approximately the same precision of the results by voivodships (Lednicki and Wesolowski, 1994). For that year 36,163 households were accepted for data processing but the target sample was 37,200 households (but 18300 households were selected additionally to get assumed size of sample). Because of financial restrictions, starting from 2001, the size of sample was decreased to the previous level.

The HBS from 2001 onwards

Starting from 2001 two sub-samples of PSU's were selected, 675 PSU's each. Theses sub-samples of primary sampling units were selected from 29,172 PSU's prepared from ESD's (from urban areas PSU's had at least 250 dwellings, and for rural areas – 150 dwellings). As for the year 2000, before selection PSU's were stratified by 16 voivodships, and in each voivodship according to the class of size of localities. Large towns constituted separate strata. Number of strata in a voivodship was from 3 to 12. Altogether 96 strata were constructed. Allocation of sample in the strata was proportional to the number of dwellings in each stratum.

PSU's were selected with probability proportional to the number of dwellings according to Hartley-Rao scheme. In each PSU 24 dwellings were selected for two years (2 dwellings for each month, and the same dwellings are surveyed in both years). Additionally, in each PSU 150 dwellings were selected independently for reserve sub sample, to be used in the case of nonresponse. Each year a new sub-sample of 675 PSU's will be selected for two years.

Weighting for HBS

Nonresponse rates in HBS are usually high, and they change considerably the socio-economic structure of households in the sample. To minimise this impact, the sample results are weighted.

First, each household in the sample is weighted in inverse proportion to the probability with which it was selected. Weighting for non-response involves the division of the sample into appropriate 'weighting classes', and within each weighting class, the weighting-up of the responding units in inverse proportion to the response rate, so as to 'make up' for the non-responding cases in that class. In the case of the HBS it is impossible to use this method since sequential selection of household for nonresponse was applied. In this case appropriate weights from external sources are used. For the HBS additionally appropriate weights from LFS (for size of households, and urban and rural relation) are applied.

Method of standard error estimation

For standard error estimation random group method was used until 2000, and starting from 2001 a method of balanced half-samples is to be used.

3.2.2. Sample design for LFS and its redesign in 1999

A sample for the LFS was selected in two stages with stratification. The strata were accepted: 49 voivodships and in each voivodship additionally several localities as sub-strata were created (one of them was rural areas). The primary sampling units were census enumeration areas (CEA's) in towns, while in rural areas PSU's were enumeration statistical districts – ESD's (in a very few cases sampling units are created by joining two or more bordered CEA's or ESD's). Dwellings were the second stage sampling units. In order to improve precision, overrepresentation of rural areas was applied, i.e. the number of dwellings sampled from the rural areas was about 10% higher than the number derived from the proportional allocation (related to the number of dwellings in population). Szarkowski and Witkowski (1994) described methodology of the LFS.

Redesign of the survey in 1999[1]

[1] The sample plan, i.e. the sample design, methods of estimation of parameters and their standard errors, was prepared by Andrzej Szarkowski, Research Center for Economic and Statistical Studies of the CSO and the Polish Academy of Sciences.

Since the fourth quarter of 1999 the LFS has been carried out as a continuous survey.

PSU's and SSU's were selected in the same way as in the previous survey, but sample allocation by 16 voivodships was changed. To get better precision of estimates by voivodships, the size of sample in voivodships was allocated nearly proportional to the square root of the number of dwellings in voivodship. Strata within voivodships were created depending on the size of localities – rural areas were included into the smallest one. Strata within voivodships were created accordingly to the situation in a particular voivodship without any fixed dividing criterion.

PSU's within strata were selected with probability proportional to the number of dwellings in a PSU. Then, a determined number of dwellings (from 4 to 9) were selected from each PSU[2].

Results of the survey are published quarterly. In every of 13 weeks in a quarter[3] interviewers visit a determined number of randomly sampled dwellings (from 1880 to 1900) and collect data concerning economic activity during the preceding week. The survey covers all people at the age 15 years or above, living in the selected dwellings. A sample of dwellings to be visited is changed every week. Weekly samples result from a random division of a quarterly sample into 13 parts. The ***quarterly sample amounts from 24,440 to 24,700 dwellings***. (GUS,2000a).

Selection of quarterly samples was performed according to the rotation pattern, which hasn't been changed since the second quarter 1993. Details are presented in the "Rotation pattern for 1999-2004", included below. The rotation pattern in the LFS can be summarised as follows[4]:

- the sample for each quarter consists of four sub-samples, currently consisting of 6110 or 6175 dwellings (according to the continuous method they are divided into 13 weekly elementary sub-samples, each consisting of 470 or 475 dwellings);
- partial rotation of sub-samples is carried out in every quarter. In a given quarter there are two sub-samples surveyed in the previous quarter, one sub-sample introduced into the survey for the first time and one sub-sample which was not surveyed in the previous quarter and was introduced into the survey exactly a year before;
- sub-samples are selected independently.

[2] In order to combine organizational requirements with the need for the high precision, a rule was coined to create strata in such a way as to select 8 dwellings in PSU's from rural areas and small towns, 6 – 7 dwellings in PSU's from medium size towns and 5 dwellings in PSU's from large cities. However, exceptions from the rule were necessary.

[3] According to Eurostat regulations, the term "Quarter" applied currently to the LFS is slightly different from the calendar quarter: every quarter in the LFS consists of 13 weeks and always starts on Monday. Thus, the first quarter of 2000 lasted from 3rd January to 3rd April.

[4] Detailed description of the sample design, estimation method and standard error calculation is to be published in *Statistics in Transition* soon.

Rotation pattern for the LFS in 1999 -2004

Sub sample No.	Years and quarters																							
	1999				2000				2001				2002				2003				2004			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1				x																				
2				x	x																			
3				-	x	x																		
4				-	-	x	x																	
5				x	-	-	x	x																
6				x	x	-	-	x	x															
7					x	x	-	-	x	x														
8						x	x	-	-	x	x													
9							x	x	-	-	x	x												
10								x	x	-	-	x	x											
11									x	x	-	-	x	x										
12										x	x	-	-	x	x									
13											x	x	-	-	x	x								
14												x	x	-	-	x	x							
15													x	x	-	-	x	x						
16														x	x	-	-	x	x					
17															x	x	-	-	x	x				
18																x	x	-	-	x	x			
19																	x	x	-	-	x	x		
20																		x	x	-	-	x	x	
21																			x	x	-	-	x	x
22																				x	x	-	-	x
23																					x	x	-	-
24																						x	x	-
25																							x	x
26																								x

Sub-sample 1, 2, 3, 4, 5 -sub-sample with shorten duration; Sub-sample 6 and further - a new sub-samples with normal duration .Source: (GUS, 2000a)

As a result of this rotation pattern, after the preliminary period, each sample is employed according to the quarterly 2-(2)-2 pattern of interviewing: two quarters in the survey, two quarters break, again two quarters in the survey and then out.

Weighting the LFS results

Weighting is performed in three stages. The purpose of steps two and three is to provide means for the nonresponse adjustment.

Step 1 Design weights

Each household in the elementary sample is weighted by the reciprocal of its selection probability. If p_i is the selection probability for dwelling i (and for all persons within), then the design weights for the elementary sample are given by:

$$\hat{w}_i' = \frac{1}{p_i} \tag{1}$$

Since the entire quarter sample consists of four (independently selected) elementary samples, basically of the same size and similar build, then we take as the design weight for the entire sample:

$$w_i' = \frac{1}{4}\hat{w}_i' \tag{2}$$

Step 2. Secondary weights

The LFS response rates are calculated in six place-of-residence categories: (1) Warsaw, (2) towns with 500 thousand up to 1 million inhabitants; (3) towns with 100 thousand up to 500 thousand inhabitants; (4) towns with 20 thousand up to 100 thousand inhabitants; (5) other towns, and (6) rural areas. (For example, in the survey conducted in the first quarter of 2000 they amounted to 0,758 for the whole country; 0,360 for Warsaw; 0,668 for towns with 500 thousand up to 1 million inhabitants; 0,658 for towns with 100 thousand up to 500 thousand inhabitants; 0,775 for towns with 20 thousand up to 100 thousand inhabitants; 0,804 for other towns and 0,887 for rural areas).

Let n_j be the weighted number of interviewed households and m_j the weighted number originally selected in weighting class j. Here, the 'weighted number' refers to the sum of step-one weights in the category one, and m_j includes only valid addresses in the sample, i.e. excludes empty and non-existent addresses, or those otherwise containing no eligible household. The non-response weight is inversely proportional to the response rate: $R_{(i)}$.

In each category j the response rate R_j is calculated according to the formula:

$$R_j = \frac{n_j}{m_j} \tag{3}$$

where n_j denotes the number of interviewed dwellings, and m_j the number of eligible addresses (we consider as ineligible the buildings destroyed or turned to non-residential use, unoccupied, occupied by foreigners, etc.).

Then the secondary weights are calculated as

$$w_{ij}'' = w_i' \frac{1}{R_j} \tag{4}$$

Final weights for the results concerning population are calculated in the third step. The purpose of this step is to adjust the LFS population estimates to the

current demographic estimates. The method is straightforward post-stratification. It involves the calculation of the weight modifiers in each of the 48 classes defined by the place-of-residence (urban/rural), sex and 12 age groups, the same as in the LFS tables. The adjustment factors are calculated by dividing the number of people in each group according to the demographic estimates[1] by the number of people in these categories calculated from the LFS results with appliance of secondary weights from the second step. Final weights result from multiplication of secondary weights by adequate modifiers.

Estimation of standard errors

Until 1999 standard errors of estimates were calculated according to the random group method. After redesigning of the LFS in the 4th quarter 1999, the Taylor linearization technique is used.

3.2.3. The sample design for the Micro-census 1995

A two-stage stratified sample design was applied (Bracha, 1996). A census enumeration district was a primary sampling unit (PSU), and a dwelling was a secondary sampling unit (SSU). The country was divided into 98 strata (49 voivodships and each voivodship was divided into urban and rural areas). For the country as a whole, the size of sample was 26246 of PSU's, i.e. 15%, and 588 thousand of dwellings, i.e. 5 %, but allocation of the sampling units into strata was done in such a way as to get nearly the same precision in each stratum. As a result of such allocation in the largest stratum sampling fraction was only 0.6%, and in the smallest – 14.3%. Selection of the sample in each stratum was self-weighting, i.e. PSU's were selected with probability proportional to the number of dwellings, and in each selected PSU the same number of dwellings was selected.

A variance of the total of characteristic Y for stratum h was estimated as follows:

$$v_h = (1 - \frac{m_h}{M_h}) \frac{F_h^2 m_h}{m_h - 1} \sum (y_{hi} - \bar{y}_h)^2 \qquad (5)$$

where y_{hi} stands for total for all dwellings for i-the PSU in stratum h,

m_h – the number of PSU's selected in stratum h,

M_h – the number of PSU's in population in stratum h,

F_h – the generalisation coefficient in stratum h,

$\bar{y}_h = \frac{1}{m_h} \sum_{i=1}^{m_h} y_{hi}$, and

[1] Due to the fact that demographic estimates relate to the whole population, including also categories not covered by the LFS i.e. people living in collective households (including soldiers living in barracks) and persons living temporary abroad, they are somewhat corrected.

$y_h = F_h m_h \bar{y}_h$ – estimate of total of Y in stratum h.

Post-enumeration survey

After the Micro-census, a post-enumeration survey was carried out (Szwalek, 1996). Out of surveyed sample for the Micro-census, 1176 census enumeration areas (CEA's) were selected, and in the selected CEA every other dwelling was chosen for checking purposes. Altogether, 12.6 thous. dwellings were checked. The survey studied coverage error, i.e. the error associated with the failure to include some population units in the frame used for sample selection (undercoverage) and the error associated with the failure to identify units represented on the frame more than once (overcoverage). (GUS, 1996)

Measurement error is characterised as the difference between the observed value of a variable and the true, but unobserved, value of that variable. As a measure of data quality an "Index of Fitting" (Szablowski et al, 1996.) was applied. This Index takes the following form:

$$IZ = 1 - \frac{s_d^2}{s_x^2 + s_y^2} \tag{6}$$

where: d = x – y

s_d^2 – stands for a variance of d,

s_x^2 – variance of variable x,

s_y^2 – variance of y.

Some results of the post-enumeration survey were published in (GUS, 1996).

6. Costs of household surveys

In any sample survey, two important questions should be answered:

(i) total cost of the survey and

(ii) the precision of the main estimates.

Both these are related to the size of the sample, given the variability of the data, the type of sampling and the method of estimation. Obviously, the larger the sample, the smaller sampling error, i.e. the greater precision of the estimates, but the higher be the cost. The survey should be so designed to provide estimates with minimum sampling errors when the total cost is fixed, or to result in the minimum total cost when the precision is assumed: a sample size fulfilling one of these conditions these conditions is called the *optimal sample size.* To obtain an optimal sample size it is necessary to construct a *cost function* and assess its parameters, variances of main estimates (Särndal et al, 1992).

Suppose that the total cost of the survey can be expressed as

$$C = c_o + \sum_{h=1}^{n} n_h c_h \tag{7}$$

where c_o is a fixed overhead cost and $c_h > 0$ is the cost of surveying one element in stratum h.

The problem of *optimum sample allocation* is formulated as the determination of the n_h that minimise the variance V subject to a fixed cost C, or, conversely, the determination on the n_h that minimise the cost C subject to a fixed precision V.

In spite of crucial importance of budgeting, cost estimation is one of the least developed aspects of survey planning. One of the problems is the often-burdensome nature of maintaining detailed cost records. Another is the difficulty of separating costs in joint endeavours, especially administrative and other indirect expenses. Nevertheless, the development and maintenance of a comprehensive cost reporting system can pay important dividends with respect to future planning and the ability to attract the necessary support of data programmes.

Here we focus our attention on assessing different cost components of a household survey, taking into account United Nations recommendations (United Nations, 1984), and results of discussions with the Polish CSO experts on a survey cost calculation. Let us start with the United Nations recommendations.

The United Nations review of major issues in national statistical organisation stresses the continuing review and analysis of costs and alternatives in setting statistical priorities (United Nations, 1984). "Ideally, priorities should be determined on the basis of analysis of costs and benefits of various alternative ways of using the scarce resources". The projected costs of a household survey programme can only be determined after the detailed plan has been drawn up.

As soon as an overall survey plan is formulated, the next steps are usually to develop some approximate budget estimates for implementing the proposals and a timetable for producing the results. Although these may be based on preliminary judgements, which are highly tentative, they have an important bearing on programme decisions and priorities.

The development of an adequate cost-reporting system is crucial because most current budgetary estimates are based on previous experience. The same practice can be found in the CSO of Poland. Even though no two operations are exactly alike and circumstances change even when the same surveys are repeated, there are usually enough similarities with prior operations to serve as a basis for reasonable current estimates. Pilot survey may be another means of developing certain cost elements.

Cost estimation must usually begin at the most detailed level of operation. For this purpose, it is obviously necessary to have a complete understanding of all the detailed steps involved. Certain specific operations are more readily measured

than others, such as (i) cost of field interviewing, (ii) travel costs, (iii) clerical coding and editing, and (iv) computer runs are. In addition to the records maintained by the personnel directly involved, it may be useful to assign cost analysts to keep detailed accounts for these kind of operations for a reasonable period, with a view to developing the necessary components of cost. The most difficult aspect, and one that is often disregarded, is the allocation of indirect and overhead costs or the cost of personnel whose responsibilities extend over several projects. For this purpose, it is necessary for administrative, professional and advisory personnel to attempt to make some realistic estimates of their time allocation among projects. Certain costs which cannot be reasonably allocated in this manner, such as for overhead personnel not otherwise budgeted, certain types of supplies, and the like, might be totalled and computed as a percentage of the overall agency budget. This percentage can be added on as an overhead cost to the sum of the direct and allocable indirect costs of each separate project.

A problem which is common in many countries, and also in the Central Statistical Office of Poland, is that certain permanent staff members are budgeted as part of overall agency costs and are not normally charged to the survey project even if assigned to it. It is important that the contributions of these members to the project be determined even if this means keeping a special set of cost records for this purpose. Otherwise, it will be impossible to ascertain the true costs of a project and planners could be misled in developing future budgets.

The Central Statistical Office of Poland has a system of household survey cost assessment, as well as other statistical surveys. It is part of general instruction for cost assessment of statistical work[1]. For each sample survey a direct cost of the survey is assessed, using previous experiences in this field, and some administrative recommendations. Such cost survey assessment includes GUS, 2001): field-interviewing c^{osts,} travel costs, material costs, services connected with the survey, incentives for increasing participation in the survey, taxes, etc. There are not included coding and editing, computer runs, methodological contributions of indirect and overhead costs or the cost of personnel whose responsibilities extend over several projects.

In table 3, as examples, costs elements ***for HBS and LFS in year 2000*** are presented.

According to accepted procedure, ***direct cost of the HBS in 2000 was 18.6 million zlotych (€ 4,567,000)*** of which interviewing costs – 14.3 million zlotych (78.2%), travel costs – 0.6 million zlotych (3.2%), incentive costs – 1.7 million zlotych (9.4%), in partly included in "material costs" and "material services".

[1] Zasady wyceny kosztów prac statystycznych realizowanych przez służby statystyki publicznej w roku 2003 (Principle of cost assessment of statistical work in official statistics in 2003). Warszawa 2001.

Taking into account, that in 2000 size of surveyed households was 36,163, it means that an average ***cost of one household was 506.5 zlotych (€126.3).***

Similar calculation was done ***for the LFS in 2000***. Total direct costs of the survey ***were 4.39 million zlotych (€1,093.2),*** of which cost of interviewing – 3.5 million zlotych (80.3%), travel costs – 186 thous. zlotych (4.2%). There is no incentive cost for LFS. Taking into account that in 2000 nearly 80 thous. households were interviewed, than ***one-interview cost 55 zlotych (€13.7).***

Table 3. Cost components of the Household Budget Survey (HBS) and the Labour Force Survey (LFS) in year 2000

Cost components	HBS		LFS	
	PZL'000	Percentage	PZL'000	Percentage
Interviewers	14,317	78.2	3,525	80.4
Travel	593	3.2	186	4.2
Material costs	919	5.0	248	5.7
Material services	1,752	9,6	255	5.8
Other services	296	1.6	75	1.7
Taxes	59	0.3	15	0.3
Other costs	381	2.1	81	1.9
Direct costs	**18,317**	**100.0**	**4,385**	**100.0**

PZL'000 – Polish Zlotych in thousands; In 2000: 1 € =4.011 zlotych.

Source: own calculation according to the GUS documentation.

These are, of course, approximate assessments of cost components of the surveys, and for statistical analysis, and mainly for using them for optimal sample allocation, different cost components should be collected. Total cost of a household survey should take into account contributions subject matter, sampling, data capture and data processing experts. Other overhead costs should be also included.

7. The Design Effects

The study of Kish and Frenkel (1974) was one of the firsts to show the effect of sampling design on the results of statistical analysis. Kish introduced the term "*design effect*" to denote the ratio of the variance of any estimate, say y, obtained from a complex design to the variance y that would apply with SRS or unrestricted sample of the same size (Kish, 1965). First he defined *Deff* as this ratio with a denominator of the SRS variance, i.e.

$$Deff = \frac{s_c^2(\bar{y})}{(1-f)\frac{s^2}{n}} \tag{8}$$

$s_c^2(\bar{y})$ denotes variance of the mean $\bar{y}$ estimated from sample for design c,

n – size of sample,

$f = \frac{n}{N}$ – sampling fraction,

N – size of population,

s^2 – sample variance ($s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$).

Next Kish (1982) introduced $Deft^2$ where the finite population correction (fpc) term (1 - *f)* is not included, i.e.

$$Deft^2 = \frac{s_c^2(\bar{y})}{\frac{s^2}{n}} \tag{9}$$

Kish (1995) gave some reasons why $Deft^2$ may be preferred to *Deff*.

Note that the above definitions relate to a mean $\bar{y}$, but they may relate to a specific survey estimate *y*, but *y* can be any estimate of interest, for instance, a mean, proportion, total or regression coefficient.

Design effects have several important uses (Kish, 1987):

1. They may be averaged for greater stability, when the computed variances are subject to greater variation, because they are based on few primary units, or "*degrees of freedom*".
2. They can and should be used to check the gross errors in variance computations. Gross errors are the most common and easiest to spot for deviations from the base of 1.
3. Their main purpose is in models and conjectures for other statistics from the same survey. But for this purpose the function $roh = deft^2/(b-1)$ is preferable, especially for subclasses.
4. They may be "borrowed" to serve in conjectures about sampling errors for other surveys.
5. They may be used for designing other surveys.

Because of budgetary and timing considerations, most household surveys are based on what are termed "cluster samples", that is, cases where the ultimate sample units are chosen in groups of various sizes within only selected parts of the country. In this type of design, the units are closer together and the costs of interviewer travel are thereby reduced. However, because units in closer

proximity, are likely to resemble each other, to some extent (that is in technical terms, they may have appreciable "inter-class correlation"), the reliability of estimates from such a sample is generally lower than that of a random sample of similar size. All multi-stage samples are cluster samples of one kind or another. There is, therefore, a trade-off between cost and reliability, whereby the "optimum" cluster size would be the one which results in the greater precision for a given cost. There are formulae in statistical texts for computing these values when the necessary information on costs and variances can be obtained or estimated (Särndal et al, 1992).

The method of analysis must take into account the survey design since it has an effect on the conclusions drawn from survey data. Many computer programs for standard statistical analyses implicitly assume independent and identically distributed observations. When the data are obtained from a sample survey, this assumption approximately holds under SRS with a small sampling fraction. But under most other sampling designs, the routine use of standard procedures can lead to erroneous variance estimates and therefore invalid conclusions.

Calculation of design effect for some characteristics may be used for d*etermination of sample size,* as given below.

The determination of sample size has to be made not only for the country as a whole but for each domain of study for which separate statistics are needed. In fact, where there are separate domains the usual procedure is to calculate the requirements for each domain. The total sample is then the sum of those separate requirements. Where no separate domains are specified, the calculation is made directly for the nation as a whole. The process of computation is similar in either instance.

Size of sample determination for cluster sampling

In determining sample size, a common procedure is to make computations using assumptions of random sampling, but then to adjust the required size upward to allow for a cluster design. Certain kinds of information or even some rough estimates for the population as a whole are needed for this purpose.

In the illustration, which follows, the computations relate to a single measure, the proportion of the unemployed to be estimated. For this illustration, it is assumed that this proportion is roughly estimated at between 10 and 20 per cent from previous studies, pre-test results or other sources. Considerable deviations from this assumption would not affect the computations very much. It is also assumed that the object is to be able to measure this proportion in the current survey within three percentage points with 0.95 confidence or within 1.96 standard errors, i.e. $\hat{p} - 1.96 \cdot \sigma_p < p < \hat{p} + 1.96 \cdot \sigma_p$.

The formula to be used in this case is:

$$\sigma_p^2 = deft^2 \frac{pq}{n}(1 - f) \tag{10}$$

where: σ_p^2 – sampling variance of p.

The required size of the sample it is obtained from (10) by the following formula:

$$n = deft^2 \frac{pq}{\sigma_p^2}(1-f) \tag{11}$$

Since it is desired here to measure p within three percentage points at 1.96 standard errors, the σ_p (standard error of p) is 0.03/1.96 = 0,0153, and thus σ^2_p = 0,000234.

p – proportion in unemployment (or about 15 per cent),

f – sampling rate (assumed 1%, then negligible)

Deft – the design effect, that is, an allowance for the difference in sampling variance between a cluster sample of the type contemplated and a random sample (it was estimated from other source, and let us assume that $deft^2$ = 1.8). Using the above estimates, we obtain n = 980 units.

The value of deft depends upon the heterogeneity (i.e. variability) of the population within the cluster with respect to the characteristic(s) being studied; the heterogeneity, in turn, usually depends upon the cluster size used. This design effect, which is based on the computed extent of "interclass correlation", may be large for characteristic such as housing, where there may be a good deal of similarity for units in the same vicinity.

Effects of clustering on sampling errors of complex analytical statistics pose problems that are more complex and more important, for several reasons. First, clustered and multistage samples are now common sources for complex statistical analyses. Second, the complexities of both the analyses and of the designs have many aspects, too many and too complex for mathematicians to develop distinct and useful distribution theories. Third, the design effects due to clustering are often both considerable and persistent, and ignoring them leads to serious overconfidence in sample results. Fourth, considerable design effects have been and are now reported widely for sampling errors of diverse analytical statistics.

For the Polish household surveys, i.e. for HBS and LFS, $deft^2$ or *deft* for several characteristics were calculated. As an exercise, and for comparisons with other countries, for several parameters *deft* were calculated for the years 2000 or 2001.

In tables 4 and 5 *deft* and relative standard errors (*CV*) for estimates of total number of unemployed from LFS are given.

For some characteristics of the HBS $deft^2$ and relative standard errors in percentage (given in parenthesis) are presented below:

- Total income..... 4.24 (1.1)

Expenditure:

- Total 4.16 (1,0)
- Food 3.53 (0.4)
- Clothing & shoes 2.72 (1.5)
- Maintenance of dwellings 4.04 (1.3)
- Health personal care 3.28 (1.7)
- Transport & communication 2.16 (4.5)
- Education 2.50 (3.9)

As may be seen from the above estimates, design effects $deft^2$ for HBS and LFS results are usually greater than 1, and for some characteristics are even greater than 4. Hence, standard errors based on simple random sample assumptions tend to underestimate the standard errors from the applied complex sample design.

Design effects are valuable in developing sample designs for new surveys. As been stressed the magnitudes of the overall design effects for key survey estimates may be used in determining the required sample size. They may be also used in analysis of collected data from complex sample designs.

Table. 4. The quarters' mean design effects ($deft^2$) and $deft^2$ and relative standard errors (CV%) for estimating total number of unemployed according to the LFS in 2000 by urban, rural, size of localities and level of education

Specification	The quarters' mean of $deft^2$		$Deft^2$ in the quarter of 2000				CV in the quarter (percentage)			
	9 quat[1]	4 quat[2]	I	II	III	IV	I	II	III	IV
Total	1.49	1.45	1.45	1.52	1.48	1.46	1.7	1.8	1.8	1.8
Urban	1.46	1.45	1.41	1.51	1.44	1.42	2.2	2.2	2.2	2.2
Rural	1.39	1.35	1.35	1.30	1.38	1.37	2.9	3.0	3.2	3.0
Towns' classes with population (population in thousands)										
500 and above	1.76	1.74	1.72	1.73	1.75	1.76	6.5	6.4	6.3	6.2
100 – 500	2.00	1.97	1.94	1.95	2.00	1.97	5.2	5.1	5.3	5.2
50 – 100	2.59	2.60	2.68	2.65	2.62	2.43	7.2	6.1	7.3	7.1
20 – 50	2.73	2.50	2.52	2.75	2.46	2.77	6.7	7.2	7.7	7.1
10 – 20	2.66	2.66	2.72	2.81	2.65	2.44	9.3	9.4	8.4	7.9
Under 10	2.88	2.66	2.84	2.49	2.51	2.79	9.3	9.3	8.9	8.5
Level of education										
Higher education	1.35	1.28	1.38	1.34	1.37	1.02	9.5	8.7	8.7	8.0
Post-secondary education	1.29	1.28	1.28	1.20	1.39	1.23	11.5	11.0	11.2	11.4
Secondary vocational education	1.25	1.24	1.27	1.26	1.24	1.19	3.6	3.6	3.6	3.4
General secondary education	1.25	1.23	1.33	1.15	1.22	1.21	5.9	5.7	5.2	5.6
Basic vocational education	1.31	1.25	1.19	1.30	1.32	1.19	2.5	2.6	2.7	2.5
Primary and incomplete education	1.48	1.48	1.50	1.49	1.48	1.46	4.1	4.0	4.2	4.1

Source: Own preparation according to calculation by Andrzej Szarkowski, Research Center for Economic and Statistical Studies of the CSO and the Polish Academy of Science

[1] The means of 9 quarters from continuous LFS from 4th quarter 1999 r. to 4th quarter 2001 r.
[2] Means of 4 quarters from the continuous LFS in 2000 .

Table. 5. The quarters' mean design effects ($deft^2$) and $deft^2$ and relative standard errors (CV%) for estimating total number of unemployed according to the LFS in 2001 r. by urban, rural, size of localities and level of education

Specification	The quarters' mean of $deft^2$ in		$Deft^2$ in the quarter of 2001				CV in the quarter (percentage)			
	2000	2001	I	II	III	IV	I	II	III	IV
Total	**1.45**	**1.54**	**1.52**	**1.58**	**1.53**	**1.53**	**1.7**	**1.7**	**1.7**	**1.7**
Urban	1.45	1.48	1.50	1.50	1.38	1.54	2.2	2.3	2.0	2.1
Rural	1.35	1.41	1.35	1.58	1.38	1.34	2.8	3.0	3.2	2.8
Towns' classes with population (population in thousands)										
500 and above	1.74	1.77	1.86	1.72	1.70	1.79	6.2	6.1	5.8	5.8
100 – 500	1.97	2.09	1.79	2.13	2.13	2.32	4.6	5.4	5.0	5.3
50 – 100	2.60	2.72	2.55	2.84	2.83	2.67	6.9	6.6	7.0	7.3
20 – 50	2.50	3.01	2.84	3.26	2.86	3.16	7.1	6.6	7.2	7.6
10 – 20	2.66	2.75	2.78	2.64	2.94	2.62	8.5	8.5	8.4	7.5
Under 10	2.66	2.89	3.13	3.55	1.63	3.23	8.5	10.1	9.2	9.3
Level of education										
Higher education	1.28	1.43	1.30	1.34	1.41	1.67	8.4	8.4	8.3	8.5
Post-secondary education	1.28	1.32	1.22	1.52	1.34	1.22	9.2	9.8	9.6	9.5
Secondary vocational education	1.24	1.29	1.30	1.34	1.30	1.20	3.5	3.5	3.5	3.3
General secondary education	1.23	1.23	1.21	1.23	1.19	1.29	5.3	5.1	5.1	5.5
Basic vocational education	1.25	1.38	1.35	1.37	1.33	1.45	2.5	2.5	2.5	2.6
Primary and incomplete education	1.48	1.50	1.37	1.57	1.58	1.49	3.8	3.9	4.2	4.1

Source: Own preparation according to calculation by Andrzej Szarkowski, Research Center for Economic and Statistical Studies of the CSO and the Polish Academy of Science.

8. Nonresponse in household surveys

Information on survey response and nonresponse can serve multiple purposes, such as:

1) informing users of data quality and the potential for nonresponse biases,
2) monitoring survey operations,
3) pointing to weak areas that need to be strengthened in future rounds of the survey,
4) evaluating data collection strategies,
5) providing measures of frame coverage,
6) analysing trends in nonresponse over time and across surveys,
7) developing methods to compensate for and to reduce nonresponse.

Nonresponse rates can be defined in several ways, depending on whether they are used to diagnose sampling activities, data collection activities or to analyse published data. For sampling requirements, nonresponse ought to be measured using sampling units. Correspondingly, for data collection activities, nonresponse ought to be based on collection units or questionnaires. The measure of nonresponse for published data can be based on the estimated contribution of nonrespondents to aggregates for key variables of interest.

Nonresponse rates should be available for predetermined breakdowns, for example by geography, types of households and size levels. If possible, the reasons for nonresponse should also be available (unable to contact, refusal categories). These can be used to produce diagnostics to establish causes of nonresponse. Nonresponse rates by interviewer and by regional office can be used as measures of operational performance. Questionnaire item nonresponse rates can be used to point to questions that need to be rethought in terms of wording or data availability.

The nonresponse issues in the Polish household surveys were widely discussed in the journal *Statistics in Transition* (Kordos, 1995, 1996). Here we would like to draw the attention to the HBS and LFS in the last decade.

Nonresponse rate for the HBS

Before presenting nonresponse rates in years 1991-2000, we would like to start with showing differences in nonresponse rates for two methods of data collection. In 1992 two methods of data collection were used: one sub-sample with the old method (three months) and one sub-sample with the new method (one-month). It is possible to compare nonresponse rates for the two methods. One can observe an impact on the nonresponse rates caused by two methods of data collection. The results are given in Table 6.

As it can be seen from table 6, nonresponse rates are smaller for about 15 percentage points for a new method when data on incomes and expenditures were recorded for one month only.

The new method was meant not only to reduce the cost of sample selection but also to bring down the non-response rates. The comparison of data from the two surveys showed that in the initial period households surveyed with the monthly rotation method were much more willing to participate in the survey than in the case of quarterly rotation (the non-response rate was only 23,2% in comparison with 38,4%).

Table 6. Nonresponse rates for the HBS using old and new methods by quarters in 1992

Length of keeping diaries	Nonresponse rates in quarters				
	Total	I	II	III	IV
	Percentages				
Three months	38.4	37.1	37.7	39.8	38.8
One month	23.2	18.0	23.1	26.6	24.6

Source: Kordos (1996, p. 1131); Zyra (1997, p. 30)

However, in the course of time, the non-response rate has shown increase, in spite of the monthly rotation method being continued. To illustrate this, in 1992 among households selected to participate in the survey for the first time, the non-response rate was 23,2%, in 1993 it was 27,6%, in 1994 a slight decrease was noted, to rise again up to over 49% in 2000.

More data concerning the non-response rates in 1992-2000 are presented in Table 7.

It can be stated that the increase in the total non-response rates was brought about by the growing rates of non-response resulting from the two mentioned below reasons:

- Refusal – and this indicator showed an increase from 10.2% in 1992 to 26.2% in 1999, and its share in the total non-response rate ranged from 44,0% to 53%.
- Not at home – in 1992 this percentage was as low as 4.5 % in 1992 but increased to 14.5% in 2000. In 2000 it was more than three times higher than in 1992.

All the other reasons affected the non-response rate insignificantly and remained approximately constant. Non-response rates are different depending from socio-economic group, size of localities, etc. For details see (Kordos, 1996).

Table 7. Households non-response in HBS in 1992-2000 by reason

Year	Total	Non-response reasons				
		refusal	temporary absent	sickness, old age	not-at home	other reasons
	percentages					
1992	23.2	10.2	3.4	3.7	4.5	1.4
1993	27.6	12.1	2.7	3.9	7.8	1.1
1994	25.3	11.3	2.5	3.9	6.9	0.7
1995	29.1	12.7	2.4	4.0	9.3	0.7
1996	31.4	13.7	2.3	4.2	10.4	0.8
1997	34.3	15.3	3.5	4.0	10.4	1.1
1998	40.7	19.5	3.7	4.5	11.9	1.1
1999	49.4	26.2	3.8	4.9	13.5	1.0
2000	49.2	25.0	4.2	4.6	14.5	0.9

Source: publications of the Central Statistical Office of Poland (GUS).

Nonresponse rate for the LFS

The LFS in Poland started in May 1992 as a quarterly sample survey. Till the third quarter 1999 the survey was conducted in the middle of each quarter, i.e. in February, May, August and November. Starting from the 4th quarter 1999, it is continuous survey, conducted as described above. Below nonresponse rates as averages of each year are given.

Table 8. Nonresponse rates in LFS in Poland in 1992-2000

Nonresponse reasons	Year								
	1992	1993	1994	1995	1996	1997	1998	1999	2000
Total.	***4.5***	***5.3***	***8.9***	***9.9***	***10.0***	***9.6***	***11.6***	***18.2***	***22.1***
Refusals	2.0	2.2	3.2	3.4	3.6	4.1	5.1	8.9	10.9

Sources: Tables of the CSO of Poland.

Nonresponse rates differ according to size of localities. The highest being in Warsaw, and the smallest in rural areas. For year 2000 weighted annual nonresponse rates by size of localities are given below:

- Warsaw – 54.5 %
- Cities (500,000 to 1 million inhab.) – 32.6%
- Cities (100,000 to 500,000 inhab.) – 33.3%
- Towns (20,000 - 100,000 inhab.) – 23.1%
- Towns (below 20,000 inhab.) – 19.0%
- Rural areas – 11.1%.

On an annual average basis, nonresponse rates in the LFS have been steadily increasing throughout most of the period 1992-2000, from 4.5% in 1992 up to 22.1% in 2000, as have refusal rates, from 2.0 % in 1992 to 10.9% in 2000.

It is evident from tables given above that nonresponse rates have increased in last years. The increase has occurred in refusals and not-at-home categories. These increases of response rates are surprisingly high in last years. It would be useful to explain reasons for such increases, and measures should be taken to keep response rate at lower levels.

9. Concluding remarks

In this paper we have presented different types of household surveys carried out by the Polish Central Statistical Office in the last decade, focussing on HBS and LFS. Sample designs and estimation methods and nonresponse rates were already presented in *Statistics in Transition* (Kordos, 1995, 1996, 1998; Szarkowski and Witkowski, 1994). For this reason only general problems in this paper are presented. Design effects and cost of household surveys were considered for the first time, and we treated them here from the point of view of international experiences and United Nations recommendations. We recommend for our statisticians to include design effects – deft – in publications together with standard errors, since both HBS and LFS are complex designs, and individual data is used in different statistical analyses. Special attention should be also paid to assessment of cost components of the household surveys.

Findings from further research shall be helpful in designing surveys and statistical programs in ways that respect respondents' concerns, which is important for receiving high levels of co-operation from the public. We believe that supervision of the fieldwork may be also important.

The Population and Housing Census conducted in 2002 is one of the most important tasks of the household surveys in the coming years. The Census will deliver not only updated sampling frames for household surveys (census enumeration areas and enumeration statistical districts) but also auxiliary information for increasing precision of estimates and for small area estimation methods which are now under study.

REFERENCES

BRACHA, Cz. (1996), Schemat losowania próby do Mikrospisu 1995, *Wiadomosci Statystyczne*, No. 3, pp.12-18.

BRACHA, Cz. (1998), Schemat losowania próby i metoda estymacji w badaniu ankietowym "stan zdrowia ludnosci" (Sample Design and Estimation Method in the Health Status of Population Survey), *Wiadomosci Statystyczne*, No. 3, pp. 9-15.

EUROSTAT (1995), *The Future of European Social Statistics: Guidelines and Strategies,* Luxembourg.

EUROSTAT (1997), *Family Budget Surveys in the EC: Methodology and Recommendations for Harmonisation*, Population and Social Conditions 3, Methods E, Luxembourg.

EUROSTAT (2001), Meeting of the Working Party Statistics on Income and Living Conditions (EU-SILC), 10-11 December 2001, Luxembourg.

GUS (1930), *Budzety rodzin robotniczych* (Budgets of Workers Families). Statystyka Polski, t. 40, Warszawa.

GUS (1970), *Teoretyczne i metodyczne problemy statystyki spolecznej* (Theoretical and Methodological Problems of Social Statistics), Biblioteka Wiadomosci Statystycznych, GUS, t. 13, Warszawa 1970.

GUS (1976), *Zagadnienia metodologiczne statystyki spoleczno-demograficznej* (Methodological Issues of Social and Demographic Statistics), Biblioteka Wiadomoœci Statystycznych, GUS, t. 26, Warszawa.

GUS (1986), *Metodyka i organizacja badań budzetów gospodarstw domowych* (Methods and Organisation of Household Budget Survey). Zeszyty Metodologiczne, No. 62, Warszawa.

GUS (1987), *Problemy integracji statystycznych badań gospodarstw domowych* (Problems of Integration of Statistical Household Surveys), Biblioteka Wiadomosci Statystycznych, t. 34, Warszawa.

*GUS (*1992), *Poverty Measurement for Economies in Transition in Eastern European Countries*, International Scientific Conference, Polish Statistical Association, CSO, Warsaw.

GUS (1996), *Indeks zgodnosci jako miara jakosci danych* (An Index of Fitting as a measure of data quality – on the basis of Post-enumeration Survey for the Microcensus 1995). *Aneks Metodyczny*, Warszawa. Pp.30.

GUS (1997a), *Stan zdrowia ludnosci Polski w 1996 r*. (Health Status of Population in Poland in 1996). Informacje i Opracowania Statystyczne, Warszawa.

GUS (1997b), *Stan zdrowia i potrzeby osób niepelnosprawnych w 1996 r*. (Health Status and Requirements of Disabled Persons in 1996). Informacje i Opracowania Statystyczne, Warszawa.

GUS (1997c), *Stan dostosowania polskiej statystyki publicznej do standardów Unii Europejskiej. Harmonogram prac dostosowawczych* (Status of Harmonisation of Polish Statistics with European Union Standards), Warszawa.

GUS (1998a), *Metodologia i organizacja mikrospisów* (Methodology and Organisation of Microcensuses), Statystyka w Praktyce, Warszawa.

GUS (1998b), *Budzet czasu ludnosci 1996* (Time Use Survey 1996), Studia i Analizy Statystyczne, Warszawa.

GUS (1998c), *Zróznicowanie warunków zycia ludnosci w Polsce w 1997 r.* (Diversification of the living conditions of population in Poland in 1997). Studia i Analizy Statystyczne. Warszawa.

GUS (1999), Metodyka badania budzetów gospodarstw domowych (Methodology of Household Budget Survey), *Zeszyty Metodyczne i Klasyfikacje),* Warszawa..

GUS (2000a), *Aktywnosc ekonomiczna ludnosci Polski – I kwartał 2000* (Labour Force Survey in Poland – I Quarter 2000). *Informacje i Opracowania Statystyczne*, Warszawa.

GUS (2000b), *Porównanie warunków zycia ludnosci we Francji i w Polsce* (Comparison of living condition in France and in Poland), *Warszawa* (in Polish and French).

GUS (2001a), *Budzety gospodarstw domowych w 2000 r*. (Household Budget Survey Results in 2000), Warszawa.

GUS (2001b), *Warunki zycia ludnosci w 2000 r.* (Living Conditions of Population in 2000), Studia i Analizy Statystyczne, Warszawa.

GUS (2001c), *Zasady wyceny kosztów prac statystycznych realizowanych przez służby statystyki publicznej w roku 2003* (Principle of cost assessment of statistical work in official statistics in 2003). Warszawa.

KORDOS, J. (1974), Reprezentacyjne badania warunków bytu ludnosci (Sample Surveys of Living Conditions). *Przeglad Statystyczny*, nr 2, pp. 195-209.

KORDOS, J. (1982), Metoda rotacyjna w badaniach budzetów rodzinnych w Polsce (Rotation Method in Household Budget Surveys in Poland). *Wiadomosci Statystyczne*, No. 9.

KORDOS, J. (1985), Towards an Integrated System of Household Surveys in Poland, *Bulletin of the International Statistical Institute,* Vol. 51, Amsterdam, Book 2 , pp. 13-18.

KORDOS, J. (1987), *Dokładność danych w badaniach społecznych* (Data Accuracy in Social Surveys), BWS, t. 35, GUS, Warszawa, pp.172.

KORDOS, J. (1988), Time Use Surveys in Poland, *Statistical Journal of the United Nations ECE*, Vol. 5, pp.159-168.

KORDOS, J. (1992), Badania panelowe dochodów i warunków życia gospodarstw domowych w krajach EWG (Panel Surveys of Household Income and Living Conditions in EU Countries). *Wiadomości Statystyczne,* nr 3, pp. 1-6.

KORDOS, J. (1995), Nonresponse Problems in Polish Household Surveys, *Statistics in Transition,* Vol. 2, No. 5, pp. 789-812.

KORDOS, J. (1996), Forty Years of the Household Budget Surveys in Poland. *Statistics in Transition*, Vol. 2, No. 7, pp.1119-1138.

KORDOS, J. (1998), Social Statistics In Poland And its Harmonisation with the European Union Standards, *Statistics in Transition*, Vol. 3, No.4, pp. 617–639

LEDNICKI, B. (1982), Schemat losowania i metoda estymacji w rotacyjnym badaniu budzetów gospodarstw domowych (Sampling Design and Estimation Method in Household Budget Rotation Survey). "*Wiadomosci Statystyczne*", No. 9.

LEDNICKI, B. and WESO£OWSKI, J. (1994), Lokalizacja próby pomiedzy subpopulacje, "*Wiadomosci Statystyczne*", No. 9, pp. 2-4.

LUSZNIEWICZ, A. (1959*), O metodach badania budzetów rodzinnych w Polsce w latach 1920-1957* (On methods of Family Budget Surveys in Poland in 1920-1957), Państwowe Wydawnictwa Gospodarcze, Warszawa.

LUSZNIEWICZ, A. (1982): *Statystyka spoleczna (*Social Statistics*)*, PWE, Warszawa.

PANEK, T. (1990), The Statistical Measurement of the Level of Living: Case Study of Poland, *Economic Notes,* No. 2, pp. 281-289.

PANEK, T. (1996), A monitoring system of the living conditions of Polish households, *Statistics in Transition,* Vol. 2, No. 6, pp. 979-1003.

SÄRNDAL, C-E., SWENSSON, B. I WRETMAN, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Varlag.

SZABLOWSKI, J., WESOLOWSKI, J. and WIECZORKOWSKI, R. (1996), Indeks zgodnoœci jako miara jakosci danych – na podstawie wyników spisu kontrolnego do Mikrospisu 1995 (Index of Fitting as a Measure of Data Quality – on basis of the Post-enumeration Survey of Microcensus 1995), "*Wiadomosci Statystyczne*", No. 4. pp. 43-49.

SZARKOWSKI, A., WITKOWSKI, J. (1994), The Polish Labour Force Survey, *Statistics in Transition*, Vol. 1, No. 4, pp. 467-483.

SZULC, S. (1967), *Metody statystyczne* (Statistical Methods), Państwowe Wydawnictwo Ekonomiczne, Warszawa.

SZWALEK, S. (1996), Spis ludnosci i mieszkań metoda reprezentacyjna (Census of Population and Housing by Sampling Method), "*Wiadomosci Statystyczne*", No. 3, pp. 1-20..

UNITED NATIONS (1984), *Handbook of Household Surveys (Revised Edition)*, Ser. F, No 31, New York.

VERMA, V. and CLEMENCEAU, A. (1996), Methodology of the European Community Household Panel, *Statistics in Transition,* Vol. 2, Nr 7, pp.1023-1062.

WOJTYNIAK, J. (1958). *Co to sa budzety rodzinne* (What are Family Budgets?). Pañstwowe Wydawnictwa Gospodarcze, Warszawa.

ZYRA, M. (1997), Prowadzenie zapisów dochodow i wydatków a odsetek uczesnictwa w badaniu budzetow (Keeping diaries of income and expenditure and nonresponse rates in household budget survey), "*Wiadomosci Statystyczne*", No. 2, pp.27-33.

STATISTICS IN TRANSITION, June 2002
Vol. 5, No. 4, pp. 591–603

THE HOUSEHOLD SAMPLE SURVEYS IN LITHUANIA

D. Krapavickaitë[1]

ABSTRACT

The current state of two main household surveys in Lithuania – the labour force survey and household budget survey – is discussed. Sampling designs, comments on nonresponse and some main estimates with comments are presented. Other household surveys are also reviewed.

Key words: Household budget survey, Labour force survey, sample survey, sample design, nonresponse,

1. Introduction

The Household Budget Survey (HBS) is not only the first household survey in Lithuania, but also the first sample survey in Lithuanian Statistics. The first HBS, which was called a Family Budget Survey (FBS) at that time, lasted a year in Lithuania (1936-1937), in which 297 families of industry and trade workers, office employees, creative workers, business men and farmers were surveyed. The families were invited to take part in the survey through newspapers, offices, and enterprises. They presented information on income, expenditure, food consumption, and average prices of goods (State committee. (1989)).

The HBS was only one regular sample survey in the statistics of a planned economy. All other sociological, demographic, agricultural and industrial studies used data of the entire population.

However, the concepts used in this survey, including income and expenditure, were different from those used in market economies. After the restitution of independence in Lithuania in 1990, the private sector started growing up, and a new sampling design had to be used to cover the whole population. The economy turned in the direction of market economy, and a new questionnaire had to be introduced in order to satisfy the new needs of the users of statistical data. The changes in the survey were started in 1992.

[1] Institute of Mathematics and Informatics, 4 Akademijos str., LT 2600 Vilnius and Lithuanian Department of Statistics, 29 Gedimino av., 2746 Vilnius, LITHUANIA.

However, the first changes made in the HBS were not sufficient. The HBS had to provide monthly estimates of average consumption and income per capita. In order to satisfy the requirements of markets, democratic institutions and Eurostat, and to provide data comparable with the data of other countries, the HBS was redesigned once more with the help of the World Bank experts, so in its last shape the HBS started in 1996, as it is described in Šniukštienė et al. (1996). The sampling design and the estimation method that uses only design weights remain unchanged up till now.

The other regular household survey is the Labour Force Survey (LFS) which started already in the transition period, in 1994. A multistage cluster sample was used at the beginning because the population register did not suit for sample selection. Many changes have been made in this survey in several years. The population register was modernised to the level suitable for sample selection in 1996. Since then the population register has been used for sample selection as a sampling frame in most cases of household surveys, including the LFS. At present the LFS questionnaire is almost compatible with the regulations of the Council of the European Union. Household selection with probabilities, proportional to their size, and calibration of design weights are used in the LFS. It is carried out twice a year - in May and November.

Other types of household surveys are conducted only once. Their number is not large. They have been carried out during the recent years by the staff of the HBS, namely, on living conditions (1997), time use (1998), elderly people (1999), household energy consumption in 1996 (1997), accessibility of health care service (1998), and on providing households with computers (2000).

2. Estimates and Errors in the Labour Force Survey

2.1. Sampling design

The population of the labour force survey are residents of Lithuania aged 15 and over.

The population register has some shortcomings: it suffers from the undercoverage and overcoverage; not all the addresses are included into the frame, part of them is wrong, and it is impossible to follow the person selected up to its right address. The population register of persons of age under the survey coverage with the addresses of residence in towns and at least the name of the village, is used as a sampling frame. This frame is actually used as a frame of addresses. The sample is constructed as follows: having selected a simple random sample of approximately 3000 persons of the sampling frame, the members of their households are also joined. Even if it turns out that, according to the address denoted, some people were not included into the list of the sample, these people living there were interviewed. The cluster sample of persons is thus obtained. All the persons living at the address selected belong to the same cluster. The actual

composition of the cluster is indicated by the interviewer when visiting the household. Sometimes there are no addresses of the citizens in the village in the population register. Then the person selected is looked for, and his/her households are included into the survey. The same situation occurs in hostels.

2.2. Sample rotation

In order to avoid great random changes in the survey results from one survey to another, the sample is not completely changed for the new survey, only one third of it is renewed. Each selected household participates in two subsequent surveys, then it omits one survey. After participating in the third survey the household is not included into subsequent surveys.

2.3. Nonresponse

The LFS is a survey of individuals, but it is impossible to speak about the nonresponse level of individuals. Sometimes the individual did not respond because he or she had changed the address but could be included into the survey by another address selected. This case cannot be considered as a real nonresponse. For this reason we mention only the nonresponse of households (addresses). A nonresponding household is considered as the address with no responding persons. Since November 1997 household nonresponse rates in LFS are presented in Table 1.

Table 1. Reasons of nonresponse of households in LFS (percents)

	Nov. 1997	**May 1998**	**Nov. 1998**	**May 1999**	**Nov. 1999**	**May 2000**	**Nov. 2000**
Refused	2.8	1.5	1.9	2.2	1.7	2.9	2.5
Not at home	4.9	5.7	2.6	3.8	3.1	3.5	3.5
Other reasons	1.9	1.7	2.8	2.6	2.9	2.8	2.7
Total	9.6	8.9	7.4	8.6	7.7	9.2	8.7

Estimates of the LFS may have some bias, due to nonresponse and inaccuracies of the population register. 2,695,851 records constituted the sampling frame (Krapavickaitė, (2000)) in May 2000. The published demographic data exceed the sampling frame by 271 230 persons, and it is not known, whether the distribution of employment in this part of the population is the same as in the registered population.

2.4. Post-stratified estimator of the total

At the first stage of this survey, from 1994 to 1996, the parameters were estimated as if it were a simple random sample. If the actual sampling design is taken into account, the distribution of the survey respondents by groups of urban/rural areas, age, and sex is slightly different from the demographic data. In order to improve this distribution the post-stratification of the sample by 12 age, 2 sex groups, and 10 counties, totally by K=240 groups, is used. In that way we want to conform the population register to demographic data, taking into consideration nonresponse and differences, and to obtain more exact estimates of employed persons. This method of estimation is discussed in J.Hörngren (1992), K.Djerf et al. (1993), D.Krapavickaitė et al. (1997) and has been used in Statistics Lithuania since 1996 until the end of 2000. Let us present this method here and introduce our notation:

M – population size according to the population register;
L – population size according to the demographic data;
K – number of post-strata;
m_i – number of registered members of the i-th household, i=1,2,...,K;
N – number of households in the population according to the population register;
n – number of responding households in the sample.

The household consisting of m_i members included into the population register who are of the age under the survey coverage has a probability $\pi_i = nm_i / M$, i=1,2,...,N to be included into the sample. The design weight of the i-th household equals $w_i = 1/\pi_i$. Every member of this household represents w_i residents belonging to the survey population.

To estimate the number of employed persons, let us denote the values of the variable y under investigation as y_{ij}=1, if the j-th person of the i-th household is employed, and y_{ij}=0, otherwise; $\delta_{ij}(k) = 1$, if a person belongs to the k-th post-stratum, and $\delta_{ij}(k) = 0$, if not, j=1,2,...,m_i, i=1,2,...,n, k=1,2,...,K. The post-stratified estimator of employed persons T in the country

$$\hat{T}_y^{(pos)} = \sum_{k=1}^{K}\sum_{i=1}^{n}\sum_{j=1}^{m_i} w_i g_k \delta_{ij}(k) y_{ij}\,, \tag{1}$$

which is approximately unbiased, is used. Its approximate variance $Var(\hat{T}_y^{(pos)})$ and the variance estimator $\hat{Var}\left(\hat{T}_y^{(pos)}\right)$ for the fixed size sample design are

$$Var(\hat{T}_y^{(pos)}) \approx \sum_{i=1}^{N}\sum_{l>i}\left[\sum_{k=1}^{K}\left(\pi_i\pi_j - \pi_{il}\right)\left(\sum_{j=1}^{m_i} w_i\delta_{ij}(k)\left(y_{ij}-\mu_k\right) - \sum_{j=1}^{m_l} w_l\delta_{lj}(k)\left(y_{lj}-\mu_k\right)\right)\right]^2, \tag{2}$$

$$\mu_k = \frac{\sum_{i=1}^{N}\sum_{j=1}^{m_i}\delta_{ij}(k)y_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{m_i}\delta_{ij}(k)},$$

$$\hat{Var}\left(\hat{T}_y^{(pos)}\right) = \frac{1-n/L}{n-1}\sum_{i=1}^{n}\sum_{l>i}\left[\sum_{k=1}^{K}\left(\sum_{j=1}^{m_i} w_i g_k \delta_{ij}(k)\left(y_{ij}-\hat{\mu}_k\right) - \sum_{j=1}^{m_l} w_l g_k \delta_{lj}(k)\left(y_{lj}-\hat{\mu}_k\right)\right)\right]^2 \quad (3)$$

$$\hat{\mu}_k = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\delta_{ij}(k)y_{ij}}{\sum_{i=1}^{n}\sum_{j=1}^{m_i}\delta_{ij}(k)}.$$

Here π_{il} in (2) and (3) is a probability that both the *i*-th and *l*-th households are included into the sample, $g_k = M_k / \hat{M}_k$, k=1,2,...,K are auxiliary post-stratification weights, M_k is the size of the *k*-th post-stratum referring to the demographic data, and

$$\hat{M}_k = \sum_{i=1}^{n}\sum_{j=1}^{m} w_i \delta_{ij}(k),\ k=1,2,...,K$$

is its estimator.

Estimator (1) is also used to estimate the number of unemployed people, however different post-strata formed by 4 age, 2 sex groups and 2 groups belonging to the Labour Exchange, by K=16 groups in total, are applied. The value of the study variable *y* equals 1, if a person is unemployed, and the value of *y* equals 0, otherwise.

The post-stratification gives more accurate estimates in comparison with the estimates that use only sampling design weights. Nevertheless, post-stratified estimators have some drawbacks. A first drawback is that the members of the same household acquire different weights, and a second one is that different weighting systems are used for the estimation of the number of employed and unemployed persons, sometimes resulting in a higher estimate of labour force than the population size in small domains.

2.5. Estimation results

Let us look through some results of the survey in November 2000. 52.5% of the respondents were women. The results of estimation are presented in Table 2. Formula (1) is used to calculate estimates, and (3) is used to estimate their variance indispensable in the calculation of the variation coefficient and design effect.

Table 2. Estimation results in LFS in November 2000

	Estimates using design weights	Estimates using post-stratification weights
Employed		
Estimate (thousands)	1420.6	1488.6
Coefficient of variation (%)	1.6	1.0
Design effect	1.7	0.80
Unemployed		
Estimate (thousands)	269.5	284.9
Coefficient of variation (%)	4.0	2.9
Design effect	1.1	0.66

As expected, the variation coefficient is lower for the estimate of employed persons because their number in the sample is larger. The results show how useful the post-stratification is as compared with the results obtained using only the sampling weights. The design effect becomes less when using the post-stratification. The main estimates of recent years and their design effects are presented in Table 3.

Table 3. Estimation results of LFS in 1998-2000

	November 1998	November 1999	May 2000	November 2000		
				Total	Women	Men
Number of employed (thousands)	1612	1559	1547	1489	745	743
Design effect of employed	0.76	0.75	0.78	0.80	0.50	0.50
Number of unemployed (thousands)	232	281	267	285	120	165
Design effect of unemployed	0.68	0.67	0.74	0.66	0.50	0.70
Unemployment rate (%)	12.6	15.2	14.7	16.1	14.3	18.9

2.6. Survey cost

The cost of one survey is about 70 000 Litas[1]. 14% of them are used for printing the questionnaires and delivering them to the respondents. 86% are used to pay interviewers for their work, to cover transport expenses of the interviewers,

[1] Exchange rate in USD fixed in 2000: 1 USD = 4 Litas.

and post office costs to deliver the filled out questionnaires to Vilnius. The expenses connected with the methodological work in sampling design and questionnaire preparation, sample selection, data entry, editing and processing are not included into this cost.

2.7. Dissemination of the results

The results of the LFS are published in the issues of Statistics Lithuania. The main results of the LFS are usually published just after the survey as a report in the monthly journal *Economic and Social Development in Lithuania, B*111. For example, Statistics Lithuania (2001 b), *Employment of population by labour force survey data* presents the main results of the survey in May 2000. Afterwards all the results are published in the special periodic issue *Labour force, employment and unemployment* (*survey data*), *C*321, May 2000. The results of the survey of recent four years are published in a single issue of Statistics Lithuania (2001 a) *Labour force, employment and unemployment* (*survey data*) 1997-2000, *C*325.

2.8. Current and expected improvements

Calibration of weights (J. – C. Deville et al. (1998)) can incorporate auxiliary information into the estimator of the total in a different way than the method of post-stratification and is void of shortcomings mentioned above, giving a unique weight for all the members of the household. In order to have a unique weighting system for the employed and unemployed, the calibration of weights, obtained by using 48 groups of cross-classified age, sex, urban/rural living place, by 51 territory group and 1 group of those who belong to the Labour Exchange, has been used since 2001 (Krapavickaitė, (2001)).

A new questionnaire was made up in 2001. 93% of its questions follow the regulations of the Council of the European Union 1575/2000. Two surveys were carried out using this questionnaire in 2001.

The LFS has to be carried out every quarter, but actually it is only twice a year because of financing problems. It is expected to carry it out four times a year in 2002.

3. Household budget survey

This survey is carried out constantly. The 10680 households sample is drawn once a year. It is divided into twelve parts and distributed for each month. The selected household participates in the survey for a month.

3.1. Sampling design

The HBS has a complex sampling design. The population of private households in Lithuania is divided into three strata, according to the type of residence. The first stratum consists of households of the largest cities: Vilnius,

Kaunas, Klaipėda, Šiauliai, and Panevėžys. The second stratum consists of the households of medium and small towns, the third stratum embraces rural households. The sample size is allocated to the strata proportionally to their size. No cost function is used for the allocation of the sample. The sampling design of these strata is as follows:

1st *stratum*: A simple random sample of 4476 persons aged 16 and over is selected from the population register in the largest cities. The sample in this stratum consists of the households of selected persons.

2nd *stratum*: The households of medium and small towns are joined into clusters according to the residence listing from 1000 to 10000 citizens.

1st stage: A random sample of 20 clusters with probabilities proportional to their size is drawn from 140 such clusters in total.

2nd stage: A simple random sample of 132 persons is drawn from each cluster, using the population register, and their households are also included into the sample.

3rd *stratum*: Village lists of households were used in the rural area until 1998. According to them the clusters consisting of 300 to 2000 households were built.

1st stage: A sample of 33 clusters with probabilities proportional to their size is drawn from the population of 463 clusters.

2nd stage: A simple random sample of 108 households was drawn in each selected cluster.

Since 1998 instead of the village household lists the population register has also been used in the rural area to draw the sample in selected clusters.

The HBS sample may be considered as if it were the sample of addresses. If it occurs that other residents live at the address selected, the household of actually living residents is surveyed. In case there are several households at the same address selected, the household of a person with the closest birthday is included into the sample. The probability for the household to be included into the sample depends on the number of members of the household included into the population register, just like in the case of LFS.

3.2. Nonresponse

As usual, nonresponse is unavoidable in the survey. Nonresponse rates starting from the beginning till 2000 are presented in Table 4.

Table 4. Reasons of nonresponse of households in HBS (percents)

	1996	1997	1998	1999	2000
Refused	10.3	12.2	15.1	14.1	13.6
Not at home	3.9	3.4	3.8	4.6	5.4

Other reasons	9.8	4.7	3.8	4.1	3.8
Total	24.0	20.3	22.7	22.8	22.8

Let us consider the nonresponse reasons in the survey of 1999, as it is described in Statistics Lithuania (2000 b). It was observed that the greatest number of nonresponding households was in the largest cities (33.1%) and the lowest nonresponse was in rural areas (14.2%). One of the main reasons was their refusal to participate in the survey. In the large cities, up to 21.9% of all the selected households refused to participate, in other towns – 10.5%, and in the rural areas – 7.2%. The reasons of their refusals are rather diverse. Some of the nonrespondents are afraid of giving information about themselves, others are not interested in the survey, still others do not trust the authorities or refuse due to the lack of time or a certain individual status. More than half of those who refused had given their consent to take part in the previous survey.

Table 5. Distribution of nonresponse in HBS in 2000 (percents)

	Total nonresponse	**Of which**		
		Not at home	**Refusals**	**Other reasons**
Total	22.8	4.6	14.2	4.1
Largest cities	33.1	7.3	21.9	3.9
Other towns	17.0	3.2	10.5	3.4
Rural area	14.2	2.2	7.2	4.8

Three strata are the same as the main response homogeneity groups. So, when estimating totals in each of the stratum, actually we reweight the sample for nonrespondents. We regard nonrespondents in the same way as responding households of the same stratum. This and the shortcomings of the population register may cause a systematic error.

3.3. Estimates and their precision

The estimates of income in Lithuania, income from employment and food consumption per capita per month in the 4th quarter of 1998, variation coefficients of these estimates and their design effects are presented in Table 6. The variance estimator of the estimator of mean in the sample of clusters drawn with replacement with probabilities proportional to their size is used in the second and third strata.

Table 6. Estimates of HBS (in Litas) and their accuracy in the 4th quarter of 1998 (per capita per month)

	Estimate	VC (%)	Design effect
Disposable income	443.3	1.95	1.41
Income from employment	236.7	3.57	1.73
Food expenditure	202.3	1.64	2.16

One can see from Table 6 that design effects of the estimates exceed one. It means that attempts have to be made in future to get more accurate estimates.

3.4. Survey cost

The total yearly cost of a survey can be indicated only approximately. Printing of the questionnaires, post office expenses to contact the households included into the sample, payment to the responding households, payment to the interviewers for their work, transport expenses of the interviewers and data entry costs are almost 900 000 Litas: 61% for payment to interviewers, 18% - taxes, 14% - payment to households, 5% for transportation, and 2% for other expenses. The expenses connected with the methodological work on sampling design and questionnaire preparation, sample selection, data editing and processing are not included into this cost.

3.5. Dissemination of the results

The results of the HBS are published in the issues of Statistics Lithuania. The main results of the HBS are published as a report in the monthly journal *Economic and Social Development in Lithuania, B*111 just after the survey of each quarter and the whole year. For example, Statistics Lithuania (2001 c), *Household income and expenditure* presents the results of the fourth quarter of 2000. Afterwards all the results are published in the special issue *Household income and expenditure* 1999, *A*350.

3.6. Expected improvements

It is expected that the nearest improvement of estimates of the HBS is to use auxiliary demographic information at the estimation stage, namely, calibration of the design weights. Thus, it is expected to get estimates with less systematic and sampling errors.

The HBS estimates on the regional level are also needed. In order to get these results, a new sampling design, that includes regions, is planned to be implemented from 2003.

4. Other household surveys

Sometimes irregular household surveys are also carried out. They are usually associated with the HBS.

The **living conditions** survey in the 3rd quarter of 1997 concerned individuals over 18 years old. The main HBS sample was used. The survey included household and individual questionnaires. The results were published in UNDP (1997).

A pilot **time use** survey was carried out in January 1997 following the methodology and questionnaire prepared by Eurostat. This survey was aimed at testing the quality of the survey methodology, evaluating the attitude of people to the surveys of this kind, and estimating the nonresponse rate. The survey data were not published, they were used only by Eurostat. The sample consisted of individuals aged 10 years and over belonging to 200 households drawn according to the stratified two – stage sample design of HBS.

The survey on **elderly people** in Lithuania was carried out in the 1st quarter of 1993. People of 60 years old and over belonging to the main HBS sample were surveyed. 1138 individuals responded. The analysis of the latest demographic data, health care and social welfare as well as the living standards of elderly people is presented in the publication "Elderly people in Lithuania. Survey results".

The survey of the **services of health care** was carried out in April – May of 1998 by the staff working in the HBS at the Centre for health law and economics. The survey population was individuals aged 16 and over. The sample of the main HBS survey was used. Individuals with the closest birthday from the household selected were included into the survey sample. The aim of the survey was to describe the health situation of population and the health care service state, when needed**.**

Statistics Lithuania conducted **the household energy 96' consumption** survey in Lithuania in March 1997 under the guidance of Eurostat in terms of methodology. Apart from Lithuania, analogous surveys have been performed in other Central and East European countries. The survey was aimed at reliable and comparable information on fuel and energy consumption in households, dwellings as well as private cars. The stratified two – stage sample of 5000 households was used for the survey. The publication "Household energy consumption in 1996" was issued.

Statistics Lithuania together with the Ministry of Public administration reforms governing and local authorities conducted a survey on **providing households with computers** in 2000. Only 5% of households in Lithuania were owners of PC's. The aim of the survey was to find out for which purposes computers were used at home and to what extent Internet was employed. The sample of the whole year HBS was used. The information leaflet about the results of the survey was published.

5. Concluding remarks

According to the provisional results of the Population and Housing Census 2001, the total Lithuanian population was 3491 thousand of constantly resident population. This amount is 202 thousand less than the demographic data show on January 1, 2001. Statistics Lithuania has now more reliable demographic data and it is expected that the systematic error will be reduced in all household surveys to be.

REFERENCES

DEVILLE J.C., SÄRNDAL C. E. (1992), Calibration Estimates in Survey Sampling. *JASA*, vol. 87, No. 418, p. 376-382.

DJERF K. (1997), Effects of Post-Stratification on the Estimates of the Finnish Labour Force survey. *Journal of Official Statistics*, vol. 13, No. 1, p. 29-39.

HÖRNGREN J. (1992*)*, *The Use of Registers as Auxiliary Information in the Swedish Labour Force Survey*. *R&D Report*. Statistics Sweden, Stockholm.

KRAPAVICKAITË D. (2000) Estimates and Errors of the Labour Force Survey. *Labour force, employment and unemployment (survey data), May* 2000. Statistics Lithuania, *C*321, pp. 8-10.

KRAPAVICKAITË D. (2001). Estimates and errors in the Lithuanian Labour Force Survey *Proceedings of the sixth International Conference. Computer data analysis and modelling. Robustness and computer intensive methods*. Minsk, v.2., p. 5-10.

KRAPAVICKAITË D., KLIMAVIČIUS G., PLIKUSAS A. (1997), On some estimators in cluster sampling. *Proceedings of the XXXVIII Conference of the Lithuanian Mathematical Society*, p. 298-303.

STATE COMMITTEE OF STATISTICS OF THE LITHUANIAN SSR (1989), *Report on the research work* "*Proposals for improvement of family budget survey*", Vilnius (in Russian).

STATISTICS LITHUANIA (1998), *Energijos suvartojimas namų ūkiuose* 1996 *metais* (*Household energy consumption in* 1996), Vilnius.

STATISTICS LITHUANIA (2000a), *Pagyvenę žmonės Lietuvoje. Tyrimo rezultatai* (*Elderly people in Lithuania. Survey results*). Vilnius.

STATISTICS LITHUANIA (2000b) *Labour force, employment and unemployment* (*survey data*) *C*321. May 2000, Vilnius.

STATISTICS LITHUANIA (2001a) *Labour force, employment and unemployment* (*survey data*) 1997-2000*,* C325. Vilnius.

STATISTICS LITHUANIA (2001b). Employment of population by labour force survey data, *Economic and Social Development in Lithuania*, *B*111. Vilnius, No 9, p. 47-55.

STATISTICS LITHUANIA (2001c). Household income and expenditure, *Economic and Social Development in Lithuania*, *B*111. Vilnius, No 1, p. 47-55.

STATISTICS LITHUANIA (2001d). *Household income and expenditure*, 2000, *A*350.Vilnius.

ŠNIUKŠTIENĖ Z., VANAGAITĖ G., BINKAUSKIENĖ G. (1996), Household budget survey in Lithuania, *Statistics in transition*, vol. 2, No 7, p. 1103-1117.

UNDP (1997), Poverty: its nature and manifestations*, Lithuanian Human Development Report. Living standards and choices.* Vilnius, p. 111-117.

STATISTICS IN TRANSITION, June 2002
Vol. 5, No. 4, pp. 605–616

THE ESTONIAN HOUSEHOLD SAMPLE SURVEYS – FOCUS ON THE LABOUR FORCE SURVEY

Merle Kurvits[1], Kaja Sõstra[2], Imbi Traat[3]

ABSTRACT

This paper describes development of the Estonian Labour Force Survey. The specific features of data collection programme, the sampling design and its changes, the nonresponse and weighting procedure are described. Some comparisons with the Household Budget Survey are made.

Keywords: design effect, nonresponse, raking-ratio, two-phase sampling, weighting

1. Introduction

Two big household surveys – the Estonian Labour Force Survey (LFS) and the Estonian Household Budget Survey (HBS) were launched by the Statistical Office of Estonia in 1995. The Household Budget Survey is described in an earlier issue of this journal by Traat, Kukk and Sõstra (2000), and will not be examined here. Only some of its operating characteristics like nonresponse rates and design effects are given in the tables for illustration and comparisons.

The paper is devoted to the development of the Labour Force Survey in Estonia. But before going to it, we would like to mention that also many other household- or individual-based surveys have been conducted by the Statistical Office of Estonia in last years. These surveys have belonged to the series of similar studies in other European countries, and the obtained information has been used for national and international comparisons. We may list here the Adults Education Survey 1997, the Time Use Survey 1999–2000, the Living Conditions Survey 1994, 1999, the Health Behaviour of Estonian Adult Population 2000.

[1] Labour Force Statistics Section, Statistical Office of Estonia, e-mail: merle@stat.vil.ee
[2] Regional Statistics Section, Statistical Office of Estonia; Institute of Mathematical Statistics, University of Tartu, e-mail: kaja@stat.vil.ee
[3] Institute of Mathematical Statistics, University of Tartu, e-mail: imbi@ut.ee
The work was supported by the grant No. 3939 of the Estonian Science Foundation

The first Labour Force Survey was conducted in 1995 and it was a one-time survey. After change in the methodology the survey became regular since 1997 but was still conducted only in one quarter of the year. Since 2000 the survey became continuous, conducted all year round. Despite of the time intervals between different survey occasions, the information collected covers time since 1989 continuously, this is due to the retrospective modulus of the survey questionnaire.

In Section 2 the history of the Estonian Labour Force Survey in 1995–1999 is briefly given. Changes in the sampling frame, sampling unit and sampling procedure are described. The attention has been paid to the data collection programme which in these years was more extensive than for an ordinary labour force survey.

The current face of the Labour Force Survey is described in Section 3. Many of its features come over from the previous survey, at the same time continuous character of the new survey has coursed a number of changes in the data collection and sampling procedures. For example, the auxiliary information for the two-phase design is obtained from two different sources, and correspondingly, different rules are used for final selection of households. The weighting procedure used for nonresponse and sex-age distribution adjustments is described. The nonresponse rates are given for all survey years. The variance estimation principles are explained. The design is characterised by the design effects of unemployment estimators. Finally, some visions for the future development of the Labour Force Survey are given.

2. Labour Force Survey in 1995-1999

The first Estonian LFS was conducted in January-April 1995 according to the methodology of the International Labour Office (ILO). The preparations for the survey started in 1993 when the Governmental Commission for Population and Social Statistics decided that the periodical labour force survey is needed in Estonia. In February 1994 the LFS Working Group was established involving scientists, representatives of the state institutions and producers of statistics. As a result the first LFS was designed and conducted.

After some break preparations for the new LFS started in 1996. In compliance with the requirements of the Statistical Office of the European Community (Eurostat), the second quarter of the year was chosen for conducting the survey. Also, household was taken to be the sampling unit (in LFS95, the person). In a sampled household all the working-age members had to be questioned. In order to ensure the continuity of time series, the questionnaire was basically the same as in LFS95. A short revision according to the requirements of Eurostat for labour force surveys in European Union was made. The questionnaire was also slightly complemented with suggestions from the Estonian ministries, higher schools and other state agencies. The new LFS was in run three years

1997-1999. During its course some modifications in the questionnaire suggested by EU standards and by the feedback from interviewers and data users were made.

2.1. Data

In years 1995-1999 the data collection programme was more extensive than for an ordinary labour force survey. The questionnaire of the LFS had a retrospective part in which the information on the preceding year(s) was asked. The aim of the expanded data collection programme in the LFS95 was to fill in the gap in the Estonian labour force data after the Population Census 1989. In the surveys of 1997–1999 the retrospective data on the shorter time intervals (since preceding survey) was collected.

The questionnaire of the LFS 1995–1999 consisted of two parts:

- the part concerning reference week which was the week preceding the interview;
- the retrospective part.

In the reference week part the employed persons were asked about the features of their main and second job – the economic activity of the enterprise/organisation, occupation, usual and actual working time. The unemployed persons were asked about the steps taken to find a job, the continuity of job seeking, the characteristics of a job they are looking for, etc.

In the retrospective part of the questionnaire the respondent's labour force state (employed, unemployed or inactive) at the previous survey time and then the changes of the state until the current survey time were recorded. The states were documented in the questionnaire by their starting and ending times (month and year), and by their characteristic features.

For the employment state (main job), the location, size, type of ownership and profit-orientation of the enterprise/organisation were recorded, but also the way of finding the job, the respondent's social status, occupation, type of employment relationship, salary, average length of the working week, existence of second jobs and interruptions of work and the reason for leaving the job.

For the unemployment state, during which the respondent did not have a job with income, but was trying to find the one and would have started to work immediately, the recorded information was – the economic status before unemployment, the steps taken to find a job, contacts with the National Labour Market Board, reception of social support (unemployment benefit, training allowance, labour market support), main sources of subsistence, and the respondent's evaluation of how he or she managed during the period of un-employment.

For periods of inactivity, during which the respondent neither worked nor sought for a job – the reason for being inactive, main sources of subsistence and

the respondent's evaluation of how he or she managed during this period were recorded.

In addition, the questionnaire retrospectively collected information about the second jobs, studies, changes of residence, housing conditions and marital status.

Due to the retrospective part of the questionnaire by only four surveys in 1995-1999, the labour force data were received for ten years, 1989-1999.

Since 2000 the LFS became continuous and the retrospective part of the questionnaire was generally dropped, only the changes in the states of the respondent (employed, unemployed, inactive) during the preceding year were recorded.

2.2. Sampling procedure

The target population of the LFS in all its runs has been the residents of Estonia between the ages of 15 and 74 on the 1st of January of the survey year. The sampling frame in 1995 was the database of the Estonian Population Census 1989. The database did not include the persons' names and addresses, they were specified separately in the Population Census Archive. Since the sampling frame was some years old, the names and addresses were checked in the address bureau. Those who died and emigrants were excluded as an overcoverage of the sampling frame. In order to approach the target number of interviews, a replacement sample was drawn (this operation was done only for 1995 survey) in which the non-respondents during the fieldwork were replaced according to the region (15 counties and Tallinn), sex, and 5-year age groups.

In the 1997-1999 surveys, instead of persons, the households were sampled and all the working-age members of the households were interviewed. The sampling frame was changed to be the Population Register from which a first-phase sample, by systematic sampling in 1997 and by stratified systematic sampling in the next surveys, was taken. For the first-phase sample the number of the working-age members of the households was specified by the co-ordinators of the interviewer network of the Statistical Office of Estonia. The Population Register, along with the help from the local authorities and Apartment and Dwelling Associations, was used for that purpose. In addition, other information was checked, like addresses, persons' names, notes about the deceased and those who had left the county.

The first-phase sample was a probability-proportional-to-size sample with inclusion probability of a household depending on the number of the working-age persons in the household. In order to achieve the equal inclusion probabilities for households (and also for working-age persons) the second-phase sample was drawn as follows – all households with one working-age member and, by systematic sampling, half of the households with two working-age members and one third of the households with three working-age members, etc., were taken into the final sample. In spite of the fixed first-phase sample size the final sample size is random. It varies in terms of households and persons over different survey

occasions. The second-phase sample size is shown in the Table 1 below. In 1995 the sampling unit was person and no two-phase sampling procedure was used. The small sample size in 1997 was caused by the low survey budget and not by random variation. Note, that due to the sampling of households the response/nonresponse exists on both the household and the person levels.

Table 1. Sample sizes and responses in 1995-1999

Survey year	Stemple size	Responded
1995	10,378 persons	9,608 persons
1997	2,845 households	2,474 households, 5,051 persons
1998	7,535 households	6,556 households, 13,090 persons
1999	7,246 households	6,310 households, 12,703 persons

3. Labour Force Survey since 2000

Since 2000 Labour Force Survey in Estonia is a continuous survey. It provides quarterly and annual results. With these features it fulfils Council Regulation 577/98 for labour force surveys in the European Union.

There were advantages and disadvantages in transition to the continuous survey. As advantages we may identify the following:

- the need to collect retrospective data disappeared;
- the results became up-to-date, published less than two months after the survey quarter;
- the information about the reference week became more thorough.

The disadvantages or difficulties were connected with the sampling procedure and with the fact that many households do not live at the addresses shown in the Population Register.

The two-phase sampling procedure assumes the number of household members to be specified in quarterly first-phase samples, to be able to draw a second-phase sample. This troublesome and time-consuming work did not suit the continuous character of the survey. The correctness of the collected first-phase information was also doubtful.

Continuous survey has put short time limits to the interviewing process. The reference weeks are spread uniformly throughout the whole year. The interview has to be conducted in the week following the reference week. Only in special cases a delay, but not more than 5 weeks, is allowed. Therefore, it was not possible any more to look for a sampled household around the Country if it did not live at the address received from the Population Register.

The solution to overcome these difficulties was to adapt the sampling methodology, developed for the Estonian Household Budget Survey in 1999. In

the HBS the auxiliary information from the Population Register is used. Accordingly, sampling of households is performed by two different rules – by address- and by person-rule (see Traat, Kukk, and Sõstra 2000). The address-rule means that the household living at the selected address is interviewed. The person-rule means that the household involving selected person must be traced and interviewed (around 15% of the sampled households). Inclusion probabilities are calculated differently for these cases.

3.1. Sampling design for LFS 2000

The sampling frame for LFS 2000 is a part of the Population Register comprising people in the age 15+ on the 1st of January of the survey year.

The sampling design is a stratified systematic two-phase sampling design. In the first phase, the information on the size of sampled households/addresses is collected. In the second phase the final sample is drawn so that households inside strata have equal inclusion probabilities. All the working-age members of each sampled household are interviewed. The working-age people are the persons between 15 and 74 on the 1st of January of the survey year.

The stratification scheme aims to produce good estimates for various subgroups of the population with efforts to get comparable estimates also on county level. The 15 counties of Estonia and the capital Tallinn are divided into four strata – Tallinn, four bigger counties, ten smaller counties, and the smallest county Hiiumaa. Different sampling rates in strata are used, the highest being for Hiiumaa. Systematic sampling procedure inside strata guarantees the proportional allocation of the sample for its counties. The result – randomly selected records from the sampling frame – is a first-phase sample.

Together with each sampled record the size of the corresponding address (frequency of the address in the frame) is received. The size of the address shows how many people in the sampling frame have this address. According to this information the sample is divided into two parts, handled by different rules:

- the address-sample – the records with complete addresses;
- the person-sample – the records with fuzzy addresses.

The complete address identifies the living place exactly, where mostly just one household lives, in rare cases more. Correspondingly, the size of this address is small. Fuzzy addresses exist in rural regions where the address is just the name of the village without any other information.

The address with size less than 7 is automatically considered as complete and corresponding record is included into the address-sample. The address with size bigger than 10 is automatically considered as fuzzy, and corresponding record is included into the person-sample. The addresses with sizes between 7 and 10 are checked manually and the decision about their inclusion is made on the basis of the given address individually. Finally, for the households in the person-

sample the number of their working-age members is specified with the help of the local governments.

The first-phase sample is a probability-proportional-to-size sample, where inclusion probability of a household in the address-sample is proportional to the size of its address, and in the person-sample to the number of its working-age members. For example, the inclusion probability of a household with two working-age members is twice as big, and with three working-age members three times as big as for the household with one working-age member.

The aim of the second-phase sampling is to yield an equal probability sample of households (and its members) inside strata. Therefore, in the person-sample all households with one working-age member and, by systematic sampling, half of the households with two working-age members, one third of the households with three working-age members, etc., are taken into the final sample. In the address-sample these operations are based on the size of the address.

In the address-sample all working-age members of the household(s) living at the address are interviewed, no matter which household lives at this address. In the person-sample similarly all working-age members of the household are interviewed, but the household has to include the sampled person, the proper household is traced within the county. Around 15% of the households are handled by person-rule.

Since 2000 the households are rotated according to a 2-2-2 rotation plan. The households are interviewed four times – during two consecutive quarters and after a two-quarter period they are again interviewed in the corresponding two quarters of the following year. According to such a rotation plan every quarter 25% of the households are participating in the survey for the first time and 50% of the households were interviewed also in the preceding quarter. In this way, there is 50% overlap between neighbouring quarters and also between the same quarters of the neighbouring years.

The effectiveness of the sampling design can be characterised by design effects. The design effect can be calculated for each estimator and it is the ratio of variances of this estimator under applied design and under simple random sampling without-replacement design. The Table 2 below gives design effects for selected estimates of two surveys – the HBS and LFS. The calculations are based on yearly data and performed by SUDAAN.

Considerably smaller design effects for HBS can be explained by the applied probability-proportional-to-household-size sampling scheme. Estimates, based on the variables approximately proportional to the household size, have smaller variability than they would have had under equal probability sampling scheme

Table 2. Design effects for surveys in 2000

	HBS	**LFS**
Total income per capita	1.34	X

Food expenditure per capita	1.19	X
Total unemployment rate	X	1.77
Total number of unemployed		1.85
Number of unemployed men	X	1.69
Number of unemployed women	X	1.63

3.2. Nonresponse

In spite of the good and effective sampling design it is never possible to get responses from all the sampled units. The nonresponse is a big problem in sample surveys. For example, the LFS sample size in 2000 was 9,127 households of which it was possible to interview 7,546 households, resulting with 15,171 working-age respondents. The nonresponse rate increases together with the response burden, being extremely high in diary based household surveys. The Table 3 gives nonresponse rates through the years for the Estonian Household Budget and Labour Force surveys.

Table 3. Nonresponse rates, in percentage

Year	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**
	HBS					
Total	44.4	50.2	44.9	46.6	47.5	35.2
	LFS					
Total	7.4	-	13.5	13.4	13.2	9.9
Not-at-home	5.4	-	9.3	9.2	9.2	4.8
Refusals	2.0	-	3.9	3.7	3.5	4.3

The main components of the total nonresponse are ‘not-at-home’ and ‘refusals’. In the HBS these components are not given in the table due to the changing classifications. But it can be said that through all these years the ‘refusals’ have formed about 50 % and ‘not-at-home’ about 25% of the total nonresponse of the HBS.

The nonresponse rate in LFS has always been much smaller than in HBS. Due to changes in the sampling design while going on to the continuous survey the overall nonresponse rate has decreased. But the rate of drop-outs classified as frame errors has somewhat increased. The new sampling procedure – sampling by address- and person-rule puts more importance to the address part of the records in the Population Register. But the register still includes addresses which are not dwellings anymore, or addresses of emigrated and dead people.

We can see from the table above that in LFS the reason ‘not-at-home’ has decreased almost twice, this is due to the application of the address-rule in the

sampling design. The rate of refusals has increased. This can be explained by time limits for the fieldwork with transition to the continuous survey, and by the loading for the households – the household participates four times in the LFS.

3.3. Weighting

The weights are formed in a sequence of steps. A weight resulting from the previous step is multiplied by the correction factor calculated at the current step. The correction factors are scaled in such a way that their sample average is unity at each step. As a result, the final weight is a product of the initial weight and correction factors. A similar weighting procedure is described by Verma (1995).

Since stratified sampling is used in LFS, the design weights are inverses of the inclusion probabilities in the strata. As a starting point, assuming equal response probabilities in strata, the initial weights are

$$w_i = \frac{N_i}{n_i}, \quad i = 1,\ldots,4,$$

where N_i is the size of the target population (persons between ages 15 and 74) and n_i is the number of respondents in stratum *i*.

Further, special calibration procedure of weights is performed to compensate possible biases in estimates, caused by the unequal response probabilities in the reality, and by other shortcomings of the survey implementation. Non-coverage of the sampling frame is also handled.

First the weights are corrected for unequal nonresponse. The weighting groups of reasonably uniform size of sampled households are formed on the basis of the place of residence of the household and of the nonresponse rate in the corresponding region. The number of weighting groups varies over quarters (3-6 in 2000).

Within each group j $(j = 1,...,k)$, the weight of an individual respondent is inversely proportional to the response rate

$$R_j = \frac{n_j}{m_j},$$

where n_j is the number of responded, and m_j the number of sampled working-age persons. As the exact number of sampled working-age persons is unknown (they come through households and due to the nonresponse at household level their number is not known), the response rate R_j has to be computed in two steps. First, the response rate at the household level is computed as

$$H_j = \frac{number\ of\ responding\ households\ in\ group\ j}{number\ of\ sampled\ households\ in\ group\ j}.$$

Then, the response rate at the individual level is computed as

$$I_j = \frac{n_j}{number\ of\ \ working-age\ \ persons\ in\ responding\ households\ in\ group\ j}.$$

Now, the response rate for individuals in group *j* is

$$R_j = H_j \times I_j . \tag{1}$$

Finally, the nonresponse correction factor in group *j* is given by

$$w_j^{(0)} = \frac{\overline{R}}{R_j},$$

where $\overline{R}$ is the overall (average) response rate,

$$\overline{R} = \frac{\sum_{j=1}^{k} n_j}{\sum_{j=1}^{k} (n_j / R_j)},$$

where R_j is given in (1).

In the next step the weights are calibrated so that they produce exact population numbers in certain subgroups. The subgroups by sex, age (5-years age groups) and the place of residence (15 counties and the capital city) are considered. For this purpose the Raking-Ratio method is applied by repeating the following two steps.

1. Let *i* denote a person in the sample, and let this person be in the weighting class *j*, defined in terms of sex/age group. Let p_j denote the proportion of the population and p'_j its estimate from the sample, calculated with corrected weights of the preceding steps. The correction factors to be applied to the weights in the first step are

$$w_i^{(1)} = \frac{p_j}{p'_j}, \quad i \in group\ \ j .$$

2. Now let *k* be the weighting class by place of residence where person *i* belongs to. Let p'_k be the estimate of the population proportion p_k (calculated with weights resulting from the preceding steps, including the first step above). The correction factors to be applied to the weights in the second step are

$$w_i^{(2)} = \frac{p_k}{p'_k}, \quad i \in group\ k .$$

This step forms a new input for step 1 and the process is repeated until the correction factors of these steps are close to 1.0. In Estonian LFS steps 1 and 2 are repeated five times.

3.4. Parameters and estimators

Most of the parameters estimated in LFS are several totals and ratios. Two basic formulae are used. The total is estimated by the weighted Horvitz-Thompson estimator

$$\hat{Y} = \sum_{i=1}^{r} w_i y_i \ ,$$

where w_i is the final weight of the respondent *i*, y_i is the value of the variable measured on the respondent *i*, and *r* denotes the number of respondents.

The ratio is estimated by the ratio of weighted Horvitz-Thompson estimators

$$\hat{R} = \frac{\hat{Y}}{\hat{X}},$$

where $\hat{X} = \sum_{i=1}^{r} w_i x_i$ with x_i being some other variable measured on the respondent *i*.

The variance estimates are calculated by SUDAAN. This software uses Taylor linearization for the ratio estimator, and can calculate variance estimates of the Horvitz-Thompson estimator under many designs. The software does not handle the exact design of the LFS, therefore the closest available design in SUDAAN is used. As a result the variance estimates are calculated under assumption of stratified with-replacement unequal-probability cluster sampling, with households as clusters. It is not so easy to evaluate the bias of variance estimates. It can be said that due to the assumption of the 'with-replacement sampling' the estimates might slightly overestimate the true variance, letting their users to stay on the safe side.

3.5. Future developments

Population Census 2000 has produced a lot of population-level information on the Estonian households and individuals. This information can be used as auxiliary data in sample surveys. Demographic data is used for calibration of weights in both the HBS and the LFS. The weighting system of these surveys will be reconsidered in connection with available Census data that reflect demographic situation in Estonia more precisely than the data used earlier. In 2002 two sets of figures – calibrated on the old and new demographic data – will be published.

Attention will be paid to the improvement of other survey phases. In 2002 the new data collection method – the computer aided telephone interviewing (CATI) with 10 laptops will be tested for LFS. There will be a trial run of face-to-face interviewing at the first contact and telephone interviewing for the three following times.

REFERENCES

TRAAT, I., KUKK, A., SÕSTRA, K. (2000) Sampling and Estimation Methods in Estonian Household Budget Survey. *Statistics in Transition*, Vol. 4, No. 6, pp. 1029-1046.

VERMA, V. (1995) Technical Report on the Labour Force Survey – Turkey. Project TUR/86/081.

STATISTICS IN TRANSITION, June 2002
Vol. 5, No. 4, pp. 617–641

HOUSEHOLD SURVEYS IN LATVIA

Janis Lapins, Edmunds Vaskis, Zaiga Priede, Signe Balina[1]

ABSTRACT

Surveys of households and individuals based on the usage of probability sampling principles have been widely implemented by the Central Statistical Bureau of Latvia since the middle of the 1990s. This paper presents a detailed analysis of the methodological and organisational principles and solutions applied to two of the surveys, the Household Budget Survey and the Labour Force Survey. The main changes that have already been made (or planned to be made) in the redesigning process of these surveys in the course of 2001 and 2002 are also discussed.

Key words: Sample survey, Household budget survey, Labour Force Survey, design effect, cost of a survey.

1. Introduction

After the 2nd World War the statistical system of Latvia was integrated into the former USSR statistical system (from 1918 to 1940 Latvia was an independent state). This means that in practice all methodological issues were solved centrally by the former *Goskomstat USSR* in Moscow delegating, with rare exceptions, to the statistical offices of the former Soviet Union republics only the executive functions. In the soviet time the method of sampling was applied rarely, mainly to the on-going Household Budget Survey and several regular surveys such as the one-off personal income and living conditions survey (conducted once in 3 to 5 years) and the time budget survey (once per 5-10 years).

[1] Dr. Math. Janis Lapins, Statistics analysis expert, Statistics department, Bank of Latvia (formerly sampling expert at Central Statistical Bureau of Latvia), K. Valdemara St. 2a, Riga, LV-1050, Latvia, e-mail janisl@bank.lv.
Mr. Edmunds Vaskis, Director of Social Statistics Department, Central Statistical Bureau of Latvia, Lacplesa St. 1, Riga LV 1301, Latvia, e-mail evaskis@csb.lv.
Ms. Zaiga Priede, Head of Labour statistics section, Central Statistical Bureau of Latvia, Lacplesa St. 1, Riga, LV 1301, Latvia, e-mail zpriede@csb.lv.
Dr. Oec. Signe Balina, Lecturer, UNIVERSITY OF LATVIA, Rainis Boulevard 19, Riga, LV-1586, Latvia, e-mail Signe.Balina@dzc.lv

Contrary to the method of territorial sampling used world-wide, the so-called "branch" sampling principles were applied to these surveys (Postnikov, 1953). This method could work only under the conditions of complete employment that was more or less guaranteed in the Soviet time. At the end of the 1970s pensioners and pensioner households were additionally selected from the administrative register.

This method collapsed when Latvia as an independent state began the process of transition from the centrally planned economy to a market economy (the independent statehood of Latvia was internationally recognised in 1991). Transition from one economic formation to another required not only fundamental alterations in the economy but also in statistics. However complicated, this was nevertheless a dynamic process. This would not have been feasible without technical assistance and training that were provided by the international organisations (ILO, Eurostat, the World Bank and others) and also without the active co-operation with the Nordic countries.

2. Implementation of up to Date Household Sampling Surveys in Latvia

The newly gained knowledge was quickly put into practice. In 1994, within the framework of the NORBALT project that was financed by the Norwegian government, preparatory work for the **Living Conditions Survey** (LCS) started (the survey was implemented in 1994) in close co-operation with the **Fafo** Institute (Institute for Applied Social Science, Oslo, Norway).

A year later, under the World Bank's technical assistance project and in collaboration with the Polish experts (team leader Prof. J. Kordos) a new continuous **Household Budget Survey** (HBS) was introduced, see (Kordos, et al, 1996). It was already during the preparatory phase that the requirement was established for the results to conform with to the EUROSTAT (1997) requirements.

Practically at the same time the new Labour Force Survey was launched (until 2001 carried out twice a year). The obtained theoretical and practical knowledge allowed us within the next years to start working independently on the preparation of several new sample surveys of households. The implementation of the Household Budget survey is described in (Lapins and Vaskis, 1996). In the following years different changes and improvements were made both in the HBS and LFS. As a rule in this paper we analyse these surveys according to the situation in 2000.

In addition to the increased theoretical capacity, a qualified interviewer service was created. It was also necessary to build up a supervisory service for household surveys. Special attention was paid to centralised training of the interviewers.

Before the introduction of the **Labour Force Survey** (LFS) in Latvia employment data were acquired primarily through estimates made on the basis of enterprise survey results on employment and the number of employees in private enterprises (peasant farms, households, private subsidiary farms including individual labour). Data obtained in this manner did not provide information about the age structure, education qualification nor was it possible to make breakdowns by employment status and occupational groups.

During the preparatory stage of the LFS in September 1995, a pilot sample survey was conducted with an objective to identify imperfections in the design of the survey questionnaire and in the instructions for its completion as well as to discover shortcomings in the organisational procedures.

The LFS was first launched in November 1995 and since then it has been conducted twice a year, in May and November. Normally data are collected by means of face-to-face interviews using paper and pencil (PAPI). In cases when respondents do not want to open the door they are asked to give an interview by phone. Mostly phone interviews are used when the same household is surveyed for the second and third time.

The results of all surveys are published in Latvian and English and they are available in printed and electronic form. Moreover, for research purposes the CBS has ensured access to the anonymous HBS, LFS and LCS micro-data files for data users in Latvia and abroad.

3. Household Budget Survey

3.1. Sampling Frames, and Sample Design and the Size of Samples

The sample of HBS is obtained using stratified two-stage probability sampling. The sample allocation between strata is made proportional to the population sizes of strata.

Sampling frame

Due to migration both within the country and to other countries, significant changes took place in the population of Latvia in the first half of the 1990s after the population census in 1989. Therefore it would be expensive and time-consuming to build the sampling frame on the basis of census data. Due to a rather specific procedure of voting in the parliament and local governments (only the citizens of Latvia have the right to vote and the lists of voters are not drawn up), it was not possible to use the lists of voters as the sampling frame either. Therefore the population register was chosen as the sampling frame in urban areas. The population register was formed in the period between 1991-1992, and it comprised more than 99% of the whole population of Latvia. In principle it was possible to use this register as the sampling frame in rural areas, too, nevertheless it was decided to use the lists of households as the sampling frame. Such lists

were compiled and continuously updated since the times of the former USSR for the purposes of agricultural and demographic statistics.

Sample desing

Households were stratified by the degree of urbanisation (cities, middle-sized towns, small towns, rural areas) and by the geographical allocation (5 regions).

The six administrative districts of Riga, the capital city, as well as 6 other largest cities formed 12 strata each consisting of only one primary sampling unit (PSU). These PSUs were included in the sample with probability 1 at the first stage of sampling.

All other towns were used as the primary sampling units in the rest of urban areas. They were distributed between 10 strata according to their size (more than 7000 inhabitants, or less than or equal to 7000 inhabitants) and location (5 geographical regions). Within each stratum PSUs were selected at the first stage of sampling with probabilities proportional to the total number of inhabitants of the PSU. In practice the so-called Sunter's method was applied (see Särndal et al (1992), p. 94-96).

At the second stage simple random sampling of persons aged 15 years and more was made. In urban areas those households were surveyed where the selected persons lived.

In rural areas households were distributed between 5 strata (5 geographical regions). PSUs were formed on the base of administrative rural territories. As a rule pagasts (civil parishes), the smallest administrative rural territories, were used as PSUs; nevertheless some of the pagasts were very small therefore they were added to the neighbouring territory.

Within each stratum PSUs were selected with probabilities proportional to the size (the number of households). At the second stage households were selected using simple random sampling.

The total sample size was equal to approximately 8000 households till the end of the 3rd quarter of 1998. Due to financial problems starting from the 4th quarter of 1998 the sample size was cut down by almost 50%, and became equal to 4104 households. The reduction of the sample size was implemented by cutting down the original sample size approximately by 50% within each urban PSU, and by surveying households within the half of the sampled rural PSUs in odd months and by surveying households within the rest of the sampled rural PSUs in even months.

In order to keep the effective sample size at the chosen level, a sequential sampling approach was applied. A refusing or non-responding household was replaced by another from the reserve list and surveyed. According to the survey procedure, households for replacement were taken in strict order. In view of the fact that the reserve list was made in the same way as the basic one, households from the reserve list had the same probability of being selected.

Thus the supervisor of fieldwork had two sampling lists at his or her disposal:

- **basic list:** Sampling list with household addresses to be surveyed in a month's time. Every month, a new list was given both to the supervisor and interviewer.
- **reserve list:** This list was given only to the supervisors. The reserve list contained households in random order. This list was used upon the decision of the supervisor in case the sampled household did not participate in the survey. In such a case the supervisor assigned a new household, keeping strictly to the sequence of the list. The reserve list was used until all households included were surveyed. Afterwards the supervisor received a new reserve list.

Calculation of inclusion probabilities and design weights

The calculation of the so-called design weights allows compensating for differences in the relative selection probabilities. For the calculation of the design weights it is necessary to determine the inclusion probabilities of the households in the sample.

In rural areas within each stratum the primary sampling units are selected with probabilities proportional to the number of households of PSU. At the second stage, simple random sampling of households is made within each selected PSU. Since sampling at two stages is made independently, the inclusion probability of (h,s,i) household (i-th household of the (h,s) PSU) in the sample in two stages can be calculated by the formula:

$$p_{hsi} = m_h \cdot \frac{H_{hs}}{H_h} \cdot \frac{n_{hs}}{H_{hs}} = m_h \cdot \frac{n_{hs}}{H_h}, \qquad (1)$$

where m_h is the number of primary sampling units selected within stratum h,

H_{hs} and H_h are the number of households in the (h,s) PSU, and the total number of households in stratum h,

n_{hs} is the number of households selected within the (h,s) PSU at the second stage.

In urban areas within each stratum the primary sampling units are selected with probabilities proportional to the number of persons of PSU. At the second stage simple random sampling of persons aged 15+ (aged 15 years and over) is made within each selected PSU. Nevertheless, a household consisting of more than one person will be surveyed if any of its members at the age of 15+ will be sampled. Thus the inclusion probability of an urban household in the sample in two stages can be calculated using formula:

$$p_{hsi} = k_{hsi} \cdot m_h \cdot \frac{N_{hs}}{N_h} \cdot \frac{n_{hs}}{N_{hs}^{15+}} \qquad (2)$$

where N_{hs}, N_{hs}^{15+} and N_h are the number of persons in the (h,s) PSU, the number of persons aged 15+ in the (h,s) PSU, and the total number of persons in stratum h respectively,

k_{hsi} is the number of household members aged 15+ within the sampled (h,s,i) household,

n_{hs} is the number of persons selected within the (h,s) PSU at the second stage.

The design weight w_{hsi} of an arbitrary household from the (h,s) PSU is calculated as inversely proportional to its inclusion probability in the sample in two stages of sampling:

$$w_{hsi} = \frac{1}{p_{hsi}}, \qquad (3)$$

where p_{hsi} is determined by formula (1) or (2), respectively.

Remarks. 1. To determine the inclusion probability of a household in the sample, it is necessary to know the number of household members aged 15+ who are registered in the Population Register. In practice the estimation of inclusion probability was based on the data on the number of household members aged 15 years and over collected during the survey.

2. The design weight (3) for a household both in rural and urban areas is adjusted according to the actual response level. Namely, within each PSU the sample size was replaced by the number of households actually surveyed within this PSU. Such approach corresponds to the assumption that within the PSU response probability is the same for all households.

3.2. Estimation methods: parameters and standard errors

The estimation procedure describes the mathematical approach to the compilation of statistics from the HBS.

Basically we are interested in the estimation of:

– population total $Y = \sum_{h=1}^{H} \sum_{s=1}^{M_h} \sum_{i=1}^{N_{hs}} y_{hsi} = \sum_{i=1}^{N} y_i$,

– domain D total $Y_D = \sum_{h=1}^{H} \sum_{s=1}^{M_h} \sum_{i=1}^{N_{hs}} (D) y_{hsi} = \sum_{i=1}^{N} (D) y_i$,

– ratio R_D of two domain totals Y_D and X_D of variables y and x $R_D = Y_D / X_D$, where

M_h is the total number of primary sampling units within stratum h ;

N_{hs} is the number of households of PSU (h,s) ;

N is the total number of households in the survey population;

y_{hsi} is the value of variable y of the (h,s,i) population household (i -th household of the s -th primary sampling unit of stratum h), y_i is the value of variable y of the i -th population household;

the index (D) of the summation symbol means that the sum of variable y is taken only over those (h,s,i) population households that belong to domain D.

An arithmetic mean $\overline{Y}_D$ of the variable y of the domain of interest D can also be written as a ratio of two totals $\overline{Y}_D = Y_D / N_D$, where N_D is the total number of households in domain D.

Remark. A formula of total (and arithmetic mean) for the whole population is a special case of the corresponding formula for a domain of interest, i.e., when the whole population is chosen as a domain of interest.

Estimating totals

The population total Y can be estimated by $\hat{Y}$,

$$\hat{Y} = \sum_{h=1}^{H} \sum_{s=1}^{m_h} \sum_{i=1}^{n_{hs}} w_{hsi} \cdot y_{hsi} = \sum_{i=1}^{n} w_i \cdot y_i \qquad (4)$$

where w_{hsi} is the weight of the (h,s,i) sampled household (w_i is the weight of the i -th sampled household).

Estimator (4) is the so-called Horvitz-Thompson estimator. It is unbiased (from the sampling point of view) which means that "on average" this estimator is equal to the population total Y. Estimator $\hat{Y}$ is also linear with respect to the measured values y_{hsji}. One can also assume that the estimator $\hat{Y}$ is

approximately normally distributed. The latter property is important if somebody is going to construct the so-called confidence interval for the population total Y.

Estimating arithmetic means, ratios and percentages

In the statistical theory it has been proven that a good estimator of the ratio $R = Y/X$ of two population totals is the ratio $\hat{R}$ of the estimates of these totals $\hat{R} = \hat{Y}/\hat{X}$.

In practice as an estimator of the ratio $R_D = Y_D/X_D$ of two domain D totals, the ratio of estimates of these domain totals is used $\hat{R}_D = \hat{Y}_D/\hat{X}_D$.

Since the arithmetic mean $\overline{Y}_D$ of the variable y of the domain of interest D can be considered as a ratio of two domain totals, as an estimator of $\overline{Y}_D$ the following one can be used $\hat{\overline{Y}}_D = \hat{Y}_D/\hat{N}_D$.

Remarks. 1. The above-mentioned domain D estimators of totals, ratios, and arithmetic mean will suit as "good" estimators of the respective parameters if in the sample there is a large number of representatives from this domain. As a rule these estimators will be of poor quality for the so-called "small domains", i.e. for those domains that have only a small number of units in the sample.

2. Although the estimators of ratios $\hat{R}$ and $\hat{R}_D$, and arithmetic means $\hat{\overline{Y}}_D$ are biased, they are widely used because their bias is usually unimportant in the large samples encountered in modern complex sample surveys.

3. Sometimes the denominator X_D of the ratio $R_D = Y_D/X_D$ is known (for example, from a population census data). In such a case it is possible to estimate this parameter using the unbiased estimator $\widetilde{R}_D = \hat{Y}_D/X_D$. Nevertheless even in this case the estimator $\hat{R}_D$ can turn out to be more precise because it can have a substantially smaller standard deviation than $\widetilde{R}_D$.

Standard deviation, confidence intervals and design effect

The value of estimators like $\hat{Y}$, $\hat{Y}_D$, $\hat{R}$, $\hat{R}_D$ and $\hat{\overline{Y}}_D$ varies depending on which units are selected in the sample.

There exist different measures of the random variation of an estimator $\hat{\theta}$ of a parameter θ. The most commonly used measure is the standard deviation $SE(\hat{\theta})$ that is defined as a square root of the variance $SE(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$. Usually it is impossible to calculate the precise value of $Var(\hat{\theta})$. However, standard deviation $SE(\hat{\theta})$ can be estimated (at least approximately) on the basis of the survey data.

As a measure of effectiveness of the sample design the so-called design effect is used. The design effect $deff(\hat{\theta})$ is defined as a ratio:

$$deff(\hat{\theta}) = \frac{\hat{V}ar(\hat{\theta})}{\hat{V}ar_{SI}(\hat{\theta})},$$

where $\hat{V}ar_{SI}(\hat{\theta})$ is the estimate of variance of $\hat{\theta}$ if the sample design used would be a simple random sampling (with the same size of sample).

Table 1. Disposable income and consumption expenditure, their standard error, and design effect (HBS, 2000)

	Estimate ± standard error (Average per household member per month, in LVL)			**Design effect (whole country)**
	Whole country	**Urban areas**	**Rural areas**	
Disposable income	69.19 ± 1.56	76.69 ± 2.10	52.13 ± 2.58	1.71
Consumption expenditure	63.84 ± 1.10	71.18 ± 1.27	47.14 ± 2.04	2.22
Food, alcoholic beverages and tobacco	25.72 ± 0.31	26.15 ± 0.34	24.76 ± 0.65	1.05
Clothing and footwear	4.24 ± 0.16	4.82 ± 0.20	2.92 ± 0.24	1.45
Housing	10.72 ± 0.24	13.23 ± 0.28	5.01 ± 0.45	1.23
Furnishing	3.14 ± 0.22	3.28 ± 0.26	2.84 ± 0.44	1.00
Health	2.66 ± 0.14	3.08 ± 0.19	1.70 ± 0.16	1.03
Transport	4.94 ± 0.25	5.75 ± 0.33	3.09 ± 0.32	1.08
Communications	3.43 ± 0.14	4.12 ± 0.18	1.88 ± 0.15	1.74
Recreation and culture	4.05 ± 0.23	4.83 ± 0.30	2.30 ± 0.29	1.53
Education	0.66 ± 0.08	0.82 ± 0.10	0.30 ± 0.07	1.00
Restaurants, cafes, hotels	1.57 ± 0.10	1.96 ± 0.13	0.69 ± 0.10	1.37
Miscellaneous goods and services	2.70 ± 0.10	3.16 ± 0.12	1.65 ± 0.14	1.19

The estimation of $Var(\hat{\theta})$, i.e. the calculation of $\hat{V}ar(\hat{\theta})$ as well as the estimation of $\hat{V}ar_{SI}(\hat{\theta})$ is a complex task therefore in practice different simplified approaches are used (see Wolter (1985)).

In the HBS variances of selected estimates for the main domains of interest (Riga and 6 largest cities, towns, and rural areas) are estimated on the basis of the so-called Taylor linearisation method using software SUDAAN. On the bases of these estimates the variances and design effects are estimated at the country level. Some of the results are presented in Table 1.

3.3. Non-response rates

The total level of non-response reached 28.7% in 2000. The non-response rates broken down by the main reasons are given in Table 2.

Table 2. Non-response rates for HBS and LFS in years 1995 to 2000

(in percentage)

	1995	1996	1997	1998	1999	2000
	HBS					
Total	…	28.7	24.1	21.9	23.1	26.1
Not-at-home	…	8.6	6.0	5.5	7.9	8.3
Refusals	…	12.0	9.2	8.5	10.1	10.3
	LFS					
Total	13.7[1]	13.3	12.4	9.8	9.4	10.1
Not-at-home	8.4	4.5	4.9	3.7	4.0	4.4
Refusals	5.3	4.4	3.9	3.0	2.8	2.5

[1] November

Different age and sex groups of persons have a different income and expenditure level and structure. In order to get unbiased estimates of income and expenditure it is important that the weighted demographic structure of surveyed households corresponds to the results of the population census or demographic statistics. Demographic data have been used for the calculation of the so-called calibration weights. The calculation was performed by means of the software product CLAN developed at Statistics Sweden.

3.4. Survey costs

The household budget survey is one of the most expensive statistical exercises. At the start of the survey its cost estimate was slightly less than 200 thousand lats (LVL). The survey cost per household was as high as 24 LVL. The resources for funding the survey were drawn from the CSB budget, and in the period between 1996-1997 financing also came from the PHARE programme funds.

One of the problems related to the high costs of the survey was the need to organise a completely new interviewer service for the implementation of a continuous survey. The main expenditure items were related to fieldwork. The

compensation of interviewers reached 44% of the total costs of the survey, followed by incentives to respondents (16%), supervisors` salaries (14%) and transportation costs (8%). The breakdown of expenditures for the HBS in 2000 is given in Table 3.

Table 3. Costs of the HBS and LFS in 2000

	HBS	LFS
Total costs of the survey in 2000 in local currency, thousand LVL	94.6	91.2
Of which (in percent)		
- wages and salaries for field staff	71.5	83.0
- incentives to respondents	7.7	-
- transportation costs	5.1	14.3
- data entry and data management	3.2	1.6
- others	7.1	1.1

Exchange rate (middle of 2000): **1 USD =0.606 LVL; 1 EUR =0.560 LVL**

In the late 1990s a discussion arouse about the necessity to build up a centralised interviewer service for all those surveys of households and individuals that are conducted by the CSB. To a great extent this is related to the wish of the survey organisers to reduce costs and to prepare for the introduction of new organisational methods into the work of interviewers. The reorganised interviewer service started operations at the beginning of January 2002.

3.5. Included modules

The HBS is a multifunctional survey with a rather broad programme. To ensure cross-sectional tabulation, a wide range of background variables is needed relating not only to household composition, household classification by various income and expenditure level groups (absolute brackets, deciles, quintiles, etc. groups) but also to the social status and employment of the economically active household members and so on. For these reasons the HBS programme included sets of questions on the employment status of household members aged 15 years and over, housing conditions and self-assessment of the current economic situation. For the most part these are permanent modules. Some modules were included for a shorter period of time.

For several years (1993-1997) a set of questions for the calculation of consumer confidence indicators was included as a separate HBS module. However, this module within the HBS was used for a short period. Starting with 1999 a separate survey on consumer confidence was launched and the samples of these two surveys did not overlap.

3.6. Incentives for increasing response rates

In any sample survey the motivation of respondents is of vital importance, the more so in the countries in transition. Apart from the public relations activities and explanation of the objectives of the survey respondents need motivation to take part in the survey on their own free will. We stopped at the proposal to pay a small amount in compensation to the households. This compensation was paid to each household after the entries in the diary (one month) were made on the day of the final interview. Beginning in September 1995 the above amount (3 LVL or approximately 4-5 USD) was paid to the reference person.

Our analysis showed that as a stimulus to participate in the survey, the paid-out amount works only with the pensioners and low-income households, especially in the countryside where the employment rate is low. The better-off households were absolutely indifferent in regard to compensation. On the whole, the compensation did not have the desired effect.

Conclusion: compensations like these have not justified themselves. It is more helpful to give a small present or a souvenir. During the survey which began in May 2001 we made presents decorated with the CSB logo. Presents of this kind allows the interviewer to use them depending on the situation within the household – either to give the present during the recruitment or at any other advantageous moment. Besides it is possible to diversify the choice of presents.

4. Labour Survey

4.1. Sample design and sample size –- use of sample rotation

The survey population includes Latvian residents starting with 15-year olds. To follow the recommendations of Eurostat, all individuals of this age who live in the same household with the sampled persons are also surveyed.

The national sample size is equal to 7,940 households. The distribution of the sample by household groups is as follows:

- Riga and 6 major cities – 2,757 households,
- other cities/towns – 1,620 households,
- rural areas – 3,563 households.

Similarly as for the HBS, the sample for urban areas is drawn from the Population Register. The sample for rural areas is based on complete household lists (since 1998 it is based on the household register developed at the Central Statistical Bureau of Latvia).

For the construction of the sample, the procedure of one or two-stage sampling is applied with stratification performed in accordance with the administrative territorial division of the country. The 6 administrative districts of Riga, the capital, and 6 major cities form 12 separate strata. Each of these 12 strata consists of only one PSU included in the sample with probability equal to 1. The population living in the remaining territories of the country is divided into 52

strata depending on whether the place of residence is a rural or urban territory and on its being under the administration of one of the 26 districts of Latvia.

Simple random sampling of persons aged 15+ is made within each selected town. All individuals aged 15 years and over living in the households of the sampled persons are surveyed, too.

Pagasts as the primary sampling units in rural areas are distributed across 26 strata depending on the respective administrative district they belong to. In 2000 all pagasts were included in the sample. Within each pagasts a sample of households of size 7 is drawn by simple random sampling. All household members within the sampled households beginning with the 15-year olds are surveyed.

According to the rotation scheme for the sample of the LFS, persons from each household are to be included in the survey three times. Within each wave of the survey, the sample replacement rate is one third of the households in every city or town or pagasts.

4.2 Estimation methods

Since persons selected from the Population Register are aged 15 years and over, households comprising a greater number of such individuals have a greater chance to be included in the sample. Therefore the design weight w_{hi} of the (h,i)–th household (i–th household of the h–th stratum consisting of only one PSU) for the city of Riga (6 districts) and for each of 6 major cities is calculated as

$$w_{hi} = \frac{N_h^{15+}}{k_{hi} \cdot n_h}, \tag{5}$$

where k_{hi} is the number of household members aged 15+ within the h–th PSU (at the same time stratum h, too);

N_h^{15+} is the number of registered population aged 15+ in the h–th PSU, and

n_h is the number of persons selected within the h–th PSU.

Of all the inhabitants of the remaining towns, approximately 77.5% live in towns included in the sample of the Labour Force Survey (within urban areas each of the 26 administrative districts the main town and at least one of the remaining towns is included in the sample). The design weight of the (h,s,i)–th household (i–th household of the s–th PSU of stratum h) is calculated as

$$w_{hsi} = \frac{N_{hs}^{15+} \cdot \sum_{s=1}^{M_h} N_{hs}^{15+}}{k_{hsi} \cdot n_{hs} \cdot \sum_{s=1}^{m_h} N_{hs}^{15+}}, \tag{6}$$

where M_h is the total number of PSU in stratum h; m_h is the number of PSU sampled within stratum h;

k_{hsi} is the number of household members aged 15+ within the (h,s) PSU;

N_{hs}^{15+} is the number of registered population aged 15+ in the (h,s) PSU, and

n_{hs} is the number of persons selected within the (h,s) PSU.

Remark. In order to determine the number of household members aged 15+ we used the survey data, i.e., the value of this indicator was fixed according to the actual position in the household on the survey day. It is noteworthy that due to several reasons these figures on the number of household members do not always coincide with the position in the Population Register at the date when the sample was obtained. Firstly, the address of some household members in the survey may differ from the address in the Population Register but sometimes the opposite may be true, that is, not all the household members registered at the given address actually live in the surveyed household. Secondly, there may have been changes in the period between the date when the sampling was made and the day of the survey. This is the reason why this approach in terms of determining the number of household members aged 15+ should be considered as approximate.

Within each pagasts the design weight of a household is calculated as a ratio of the total number of households in the pagasts, and the sample size in the pagasts:

$$w_{hsi} = \frac{H_{hs}}{n_{hs}}. \tag{7}$$

Remarks. 1. In the same way as for the HBS the design weights both in rural and urban areas are adjusted according to the actual response level within each PSU.

2. Weights (5) – (7) for urban and rural population in the survey were adjusted on the basis of urban and rural population data broken down by 26 age and sex groups, in order to make the distribution of urban and rural data across the age and sex groups in the sample and in Latvia consistent.

Along with every person included in the sample during the survey, other individuals of the respective age (who live in the same household with the person under survey) are surveyed, too. Therefore all persons aged 15+ from the same household have equal probabilities to be surveyed.

Similarly as for the HBS, the Horvitz-Thompson estimates are used for estimation purposes in the LFS, too.

Using software SUDAAN sampling errors and design effects were estimated for two parameters of the LFS of 2000:

- the total unemployment rate (ratio of the total number of jobseekers and the total number of economically active persons), and
- the percentage of women among jobseekers (ratio of the total number of women-jobseekers to the total number of jobseekers of both sexes).

Since approximately 66% of the households surveyed in May participated in the survey also in November, these estimates were obtained separately for each of the two survey waves (May 2000 and November 2000). These estimates are summarised in Table 4.

Table 4. Estimated design effect and sampling error for LFS in 2000

	May 2000			**November 2000**		
	Estimate	**SE**	**Deff**	**Estimate**	**SE**	**Deff**
Total unemployment rate	0.1414	0.0048	1.60	0.1440	0.0047	1.57
Percentage of women among jobseekers	0.4667	0.0165	1.32	0.4725	0.0166	1.38

4.3. Non-response rates

One of the activities for increasing the response rate is the proxy interview. According to the results of the survey about $^1/_3$ of interviews are proxy interviews. The non-response rates in the LFS in 1995-2000 are given in Table 2.

Since 1998 according to the co-operation contract between the CSB and the Ministry of Agriculture the organisers and consultants of farm work of the Ministry of Agriculture were engaged as interviewers in rural territories. It allowed substantially increasing the sample size in rural areas as well as distributing it over all pagasts. The non-response rate in rural areas is substantially lower than in urban areas (the percentage of refusals in rural areas is particularly small – only about 0.5%. This is the main reason why the non-response rate (and particularly the percentage of refusals) decreased starting from 1998.

4.4. Survey costs

Total costs of the LFS in 2000 (for two waves) were equal to 91,2 thousand LVL. The main component of costs is wages and salaries of the field staff - 75,0

thousand LVL (82% of the total costs). Payment of interviewers depends on the number of successful interviews (filled-in questionnaires). For the first interview 1,70 LVL (before taxes) are paid, and for the second and third - 1,30 LVL. The breakdown of expenditures for the LFS in 2000 is given in Table 3.

4.5. Included modules

In co-operation with the Academic Information Centre of the Ministry of Education and Science a supplementary questionnaire (module) was designed for the survey in November 1997. The main purpose for this was to get information on the effectiveness of vocational education according to the labour market demands.

According to Commission Regulation 1924/1999 and Commission Regulation 1925/1999 in May 2000 LFS was supplemented with an ad hoc module on transition from school to working life. In May 2000 by the request of the Academic Information Centre a set of questions on the effectiveness of vocational education was repeated, too.

5. Redesigning of HBS and LFS in 2001 – 2002

5.1. Necessity of redesigning the HBS and LFS. Expected improvements

Our experience of running a permanent HBS survey in the course of five years has given us quite a number of new ideas. In these years the survey programme was altered although the effect from these changes was not always sufficient. In 2000 we started to work intensively to modify the survey design. The reasons behind the changes were as follows:

- cluster effect of surveyed households in rural areas,
- high level of non-response,
- uneven workload of the interviewers,
- measurement problems concerning the purchase of durable and low frequency income and expenditure items.

Another measurement problem was related to the income of farmers' households, as well as the expenditures of city dwellers` households for space heating, both having a strong seasonality. Since according to the design of the HBS survey instruments, information on each sampled household was collected only for one-month period, many households can be wrongly classified by deciles of their income as well as by their expenditure per capita.

5.2. Development of the new sample design

Last redesigning of the sample design for HBS was made on the basis of the population census, which was carried out in spring 2000. Starting in January 2002

significant changes in the sample design are made for the LFS – this survey now is carried out as a continuous one. For both surveys the same interviewer network is used. Therefore starting from January 2002 the samples of both surveys – the HBS and LFS are co-ordinated.

For both surveys households are stratified only by the degree of urbanisation (Riga, the capital city; 6 other largest cities; towns; rural areas). As a rule population census enumeration areas are used as the primary sampling units in both urban and rural territories. There is a small number of mixed population census enumeration areas. On the bases of each mixed census enumeration area two PSUs are formed – one is the urban PSU containing only urban population and the other is the rural PSU containing only rural population.

Sample size and sample allocation

For both surveys, the HBS and LFS, the annual household sample is evenly distributed over time (the same number of households participates in the survey within each of the 52 weeks of the year). The developed sampling procedures guarantee that within each quarter the sample of PSUs is evenly distributed over territories, too. At the same time the total annual transportation expenses of interviewers are kept at the previous level. From one to the next week the sampled PSUs are completely rotated. Within each stratum the sample of PSUs can be subdivided into three equally sized sub-samples: one of them contains PSUs included in the sample for the first time, another contains PSUs included in the sample for the second time, and the last one contains PSUs included in the sample for the third time. Altogether within each selected PSU the sample of households is taken 3 times with a 26 weeks' time shift. It offers an opportunity to keep each selected household in the sample of the LFS exactly 3 times with 6 months time shift as it was done in the LFS till the end of 2001. At the same time it is possible to sample new households for the HBS anytime when some PSU is used repeatedly.

Within each week the total number of the sampled PSUs for the LFS is equal to 24 (6 in Riga, 6 in other large cities, 6 in towns, and 6 in rural areas). The number of households selected within one PSU is equal to 10 in rural areas, 11 – in Riga, 6 – in the other 6 largest cities, and 6 in towns.

The sample for the HBS is taken within the same 6 PSUs in Riga and rural areas, as well as in the three of the 6 sampled PSUs in large cities and towns. Within one PSU the number of selected households for HBS is equal to 7 (or 8) in Riga and in large cities, it is equal to 6 (or 7) within each of the 3 PSUs in towns, and equal to 5 within each PSU in rural areas. The samples for LFS and HBS surveys do not overlap. All three times (when the same PSU is selected with a time shift equal to 6 months) the same sample of households is used for the LFS, whereas for the HBS sample of households it is always completely renovated. Comparative sample allocation for HBS and LFS in 2000 and in 2002 is given in Tables 5 and 6.

Table 5. Annual sample size for HBS in 2000 and in 2002

	2000	2002 (planned)
Number of sampling units		
total number of different primary sampling units	23 in urban areas, 29 in rural areas	416 in urban areas, 208 in rural areas
total number of secondary sampling units taken from:	4104	4472(urban) + 1560 (rural)
the basic sampling lists	1292	-
taken from the reserve sampling lists		
total number of households in the sample	5396	6032

Table 6. Annual sample size for LFS in 2000 and in 2002

	2000	2002 (planned)
Number of sampling units		
total number of different primary sampling units	44 in urban areas, 509 in rural areas	624 in urban areas, 208 in rural areas
total number of secondary sampling units	15880 (5294 surveyed once, and 5293 surveyed both in May and November)	10296 (3432 surveyed once, and 3432 surveyed twice with 6 months time interval)
total number of households in the sample	15880 (8754 in urban areas; 7126in rural areas)	10296 (7176 in urban areas, 3120 in rural areas)

Sampling of PSU

Primary sampling units are selected within each stratum using systematic probability proportional to size sampling with a random starting point.

Within each stratum the selection of primary sampling units is made in the following steps:

- The primary sampling units are listed by geographical region, and within a geographical region in a serpentine order that places units containing similar types of people together.
- Cumulative totals of the size measure (the number of households estimated from census data) of the primary sampling units are formed throughout each stratum.
- For each stratum h the number $\Delta_h = \dfrac{11 \cdot \hat{H}_h + c}{78}$ (used as a sampling step) is calculated. $\hat{H}_h$ is the estimated total number of households of stratum h;

c is a constant larger then the maximum size of PSUs that ensures the rotation of the sampled PSU after it has been included in the sample for 3 times.

- For each stratum a random number ξ_h uniformly distributed within the interval $[0, 1]$ is chosen, and the following six starting points are calculated:

$$a_{h1} = \hat{H}_h \cdot \xi_h,\ a_{h2} = \hat{H}_h \cdot \xi_h + 26 \cdot \Delta_h \ (\text{mod}\ \hat{H}_h),$$

$$a_{h3} = \hat{H}_h \cdot \xi_h + 52 \cdot \Delta_h \ (\text{mod}\ \hat{H}_h),\ a_{h4} = \hat{H}_h \cdot (\xi_h + \frac{1}{2}) \ (\text{mod}\ \hat{H}_h),$$

$$a_{h5} = \hat{H}_h \cdot (\xi_h + \frac{1}{2}) + 26 \cdot \Delta_h \ (\text{mod}\ \hat{H}_h),$$

$$a_{h6} = \hat{H}_h \cdot (\xi_h + \frac{1}{2}) + 52 \cdot \Delta_h \ (\text{mod}\ \hat{H}_h).$$

- Within stratum h for the j-th week ($j = 1, 2, ...$) those 6 primary sampling units are selected whose cumulative totals are the first in the list of cumulative totals exceeding the numbers $a_{hi} + \Delta_h \cdot (j - 1)$, $i = 1, 2, ..., 6$.

In the second stage households from each sampled PSU are selected by simple random sampling.

5.3. Development of new survey instruments

In 1996, in collaboration with the World Bank and in the framework of an experimental Recall Survey the Central Statistical Bureau attempted to compare the precision of the measuring methods of diary entries and a retrospective survey. Unfortunately the analysis of the survey results was not completed. Nevertheless, data analysis performed by the Central Statistical Bureau using rather simple methods demonstrated that the diary method was quite good in regard to regular purchases and payments. Concerning the durable goods, the recall approach witnessed considerably higher measurement precision for rarely made purchases and payments. This problem concerns not only the household's expenditure. Similar problems also crop up in regard of the income (seasonal income, irregular social benefits, etc.). The retrospective method is actually applied to **HBS** in many countries. As a rule without this method the sampling error of estimated expenditure for rare items would be high. All these arguments convinced us of the necessity to introduce the retrospective method into the HBS.

Prior to taking a decision, there were discussions about the length of the retrospective survey periods that had to be introduced. We chose a unified retrospective period of 12 months for the following income and expenditure categories of the survey:

- durable goods and more rarely made purchases and payments (55 items),
- seasonal income (2 items),
- fringe benefits from the employer and social benefits in kind,

- revenue and expenditure in cash from agricultural production activities in the household, etc.

We arrived at a conclusion that the diary of consumption expenditures must be more practical and convenient for daily recording.

We made a conclusion that in the diary a separate page or full opening should be devoted to each day. This helps the household to be better oriented in time and to follow up their entries. Following discussions about the length of the intensive period of filling-in the diary we nevertheless decided to keep the 2-week period. We are fully aware that this means an additional workload to the respondent but under the conditions of transition economy one week's intensive recording is too short a period and would not ensure accuracy of records. During the next two weeks records in the diary were made only of non-food goods in the form of register.

The new diary did not cover questions about revenues and outgoings in the household from agricultural production activities (APA). The reason behind this is the fact that recording of these items on the diary basis leads to excessive dominance of seasonal factors in net revenues from APA. Households participating in the survey in spring record a low-income level (quite frequently even negative) which although rather high in the harvest and post-harvest season distorts the real picture. In this way a household that at the beginning of the season invests in agricultural production can be classified as an indigent household. This, taking into account its income at the end of the season, can be a misleading assessment.

To reach full compliance with the EU requirements, the LFS questionnaire was redesigned in 2001. All previously missing variables on education and training, the situation a year ago and some characteristics of the first job are included in the new questionnaire. According to the requirements changes in periodicity of the survey (from twice a year to a continuous survey) were made as well. The new questionnaire was tested during the pilot survey in May 2001. Training for the interviewers took place in December 2001. The continuous LFS started in January 2002.

After publishing of the final results of the Population Census 2000 a decision should be made about the revision of LFS data. The main reason for this is the difference in the population size in comparison with the previous estimates.

5.4. Developing of new interviewer network and organisation of fieldwork operations

It is believed that in the household survey interviewers are the most valuable persons in the survey. Their qualification and motivation to work play an exceedingly important role in the successful implementation of the survey.

The existing until December 2000 HBS interviewer service was one of the best-administered and supervised services although it had several shortcomings:

- due to the reduced sample size in 1998 the workload of the interviewers, especially in the pagasts (civil parishes), is small explaining why the interviewers quite often tend to look for another job;
- the interviewers lived and performed field work only in those territories (towns and pagasts), which in 1995 were included in the sample. It was impossible to involve them in other surveys conducted outside the territories they lived in.
- changes of the sampled territories required recruitment and training of new interviewers.

Due to the above mentioned reasons the possibilities of the interviewer service were not used in full thus making fieldwork more expensive.

From the organisational viewpoint, it is necessary to ensure strong motivation in the work of the interviewers and sustainability of the interviewer service, its ability to adapt to the working conditions that are sometimes prone to dynamic changes. It is feasible only within the framework of a centralised interviewer service that takes upon itself the implementation of various household and individual surveys.

Until now it has not been possible to create such an interviewers` service as there was only one continuous survey with a rather small sample size but the LFS having a large sample was carried out in waves twice a year. Taking into account that a continuous LFS would start in January 2002, preconditions were created at least for balancing the work portfolio of the interviewers. The objective is to set up such a service that would:

- guarantee the collection of survey data necessary for producing social statistics according to unified standards;
- provide a complex approach to the sample surveys of households and individuals;
- improve the co-ordination and supervision of fieldwork;
- allow better utilisation of the existing human resources;
- create a possibility of more efficient utilisation of the financial resources that are included in the cost estimate of the survey;
- create preconditions for the implementation of commercial surveys that are relevant for the sustainability of statistical programmes.

Previous interviewers in rural areas were recruited from the local inhabitants. Under the conditions of the new interviewer service, interviewers have to be mobile. This means the availability of a car so that the interviewer might work within several administrative territories. This has already been secured for the new HBS interviewer service, which started functioning in May 2001. The reorganised (centralised) interviewer service of the CSB started operations at the beginning of January 2002.

Work is being done to introduce the CAPI method in the interviews. This method would considerably simplify the data “cleaning” procedures and speed up data transmission to headquarters.

5.5. First evaluation of different aspects of the new Household budget survey

It is still somewhat early to judge the results of the new Household budget survey as only 8-9 months have passed since it started. However, some preliminary conclusions can be drawn. Among the positive results there is the improved territorial coverage of the sample, the workload of the interviewers has become more even and the use of transportation costs is more economical. Centralised coding of the diaries in headquarters has revealed some problems in the training of the interviewers. Previously when the data entry of the diary took place in the regional statistical units, many registration errors were not discovered. The advantages of the centralisation of this work have become obvious. Currently we have intensified work with the interviewers so that in the future errors should be eliminated while the interviewer is still with the household. We are becoming increasingly convinced that it is necessary to work more frequently and efficiently just with the interviewers and that the supervisory stage as the mediator between the headquarters and the interviewers is not always efficient enough. The main conclusion: it is necessary in the future to organise the work of the interviewers from the headquarters, combining it with supervisory trips to the interviewers, without overtly bureaucratising the supervisory functions.

At the same time, the new questionnaires have shown their positive and also a few negative sides. Regrettably, the design of questionnaires requires improvements in the future. In spite of the improvements, the diaries are too sophisticated and complicated for the respondents to understand them fully. It is necessary to minimise the range of the recall items concerning the consumption expenditure. This means that the work on improving the questionnaire design will continue.

6. Other Household Surveys

The CSB of Latvia in co-operation with the **Fafo** Institute conducted the Living Conditions Survey as part of the NORBALT and NORBALT II projects twice, in 1994 and in 1999. The NORBALT surveys give updated and policy-relevant information on the living conditions in the Baltic countries, with special focus on poverty and economic resources, housing and residential environment, education, the labour market, social integration and health. The Living Conditions Surveys were conducted concurrently in three Baltic countries during the autumn of 1994 (NORBALT I) and 1999 (NORBALT II). The NORBALT II project was funded by the Norwegian Ministry of Foreign Affairs, the Nordic Council of Ministers, the Norwegian Research Council and the Governments of the Baltic States.

As the questions in both the 1994 and 1999 surveys in all three countries were comparable, this facilitated comparative analyses both across time and

countries. Moreover, as the surveys were designed after the same model as the Scandinavian living condition surveys, comparative analyses may also include the Scandinavian countries. The first results of the last survey appeared in the publication Living Conditions in Latvia (2001).

In 1999, during the LCS 3081 households were interviewed in Latvia. The non-response rate in the LCS was quite low: 10.5% in 1994 and 13.0% in 1999.

Several other household surveys were initiated in the second half of the 1990s. Among them are the Family and Fertility Survey (1995), the Time Use Survey with a small sample (1996) and the Consumer Confidence Survey (1993-1999). Besides efforts were also directed to the preparation of other surveys (the Domestic Tourism Survey, the Health Interview Survey, the Home and Leisure Accidents Survey and others).

Starting from 1996 the Traveller Border Survey is conducted three or four times per year. Both traveller flows – Latvian residents returning from abroad and foreign travellers leaving Latvia are surveyed.

In the second half of the 1990s the CSB of Latvia made considerable progress in the implementation of different business sample surveys. An overview of the methodology used in the annual business structure survey, the monthly retail trade enterprise survey and the survey of transport of goods by road is presented in (Lapins, 1997).

7. Concluding Remarks

The HBS and the LFS provide relevant information on social processes in the country, the situation in households as well as in the labour market both to the policy makers for decision taking and to the general public. The experience with the HBS was very important. It helped us to understand how to improve the survey programme and sampling procedure and how to organise fieldwork. All that allowed us to start not only the re-designed but in many respects a new HBS in May 2001. For all this we must thank our consultants – professors from Poland, Sweden and Denmark and also extend our gratitude to The World Bank.

In 2000 the LFS in Latvia was almost compliant with EU requirements. Full redesign of the questionnaire for the LFS was made in 2001 and full compliance with EU requirements is reached. A continuous LFS started from January 2002.

Now Latvia is on the way to become a European Union country. It obliges the CSB of Latvia to develop the system of production of official statistics in full compliance with the requirements of Eurostat. The redesigned LFS will be carried out according to the Commission regulations in the field of labour statistics. At the same time the existing until now deviations from the probability sample design will be eliminated, too.

We hope that the redesigning of HBS and LFS will substantially improve the quality of survey data. We also hope that the co-ordination of samples of the two main household surveys, the HBS and LFS, will promote more effective use of

survey resources. The development of the new highly professional and mobile interviewer service will make planning and execution of the new sample surveys and different ad-hoc surveys more flexible.

The CSB is also planning to introduce modern data collection methodologies. As one of the first steps planned in this direction is to implement the CAPI technology within the next couple of years.

Acknowledgements

The authors are grateful to the referees for their constructive comments and suggestions, which improved upon an earlier version of this paper. The authors are also grateful to the Ms. Zane Ozola for her help in improving English.

REFERENCES

EUROSTAT (1997), Family Budget Surveys in the EC: Methodology and Recommendations for Harmonisation: Theme: Population and social conditions 3. Methods E. Brussels, Luxembourg.

KORDOS, J., LAGODZINSKI, W., OCHOCKI, A., ŻYRA, M. (1996): The World Bank Project. Redesigning of Household Survey. Final report. Warsaw.

LAPINS, J. (1997): Sampling Surveys in Latvia: Current Situation Problems and Future Developments, STATISTICS IN TRANSITION, The Journal of the Polish Statistical Association, Volume 3, Number 2, p. 281-292.

LAPINS, J., Vaskis, E. (1996): The New Household Budget Survey in Latvia, STATISTICS IN TRANSITION, The Journal of the Polish Statistical Association, Volume 2, Number 7, p. 1103-1118.

Living Conditions in Latvia. An Analytical Report (2001); Central Statistical Bureau of Latvia, Riga.

POSTNIKOV, S. (1953): *O metodach otbora semej rabočich služaščich i kolchoznikov dla obsledovanjia ich budžeta* (On methods of selection of families workers, non-workers employees and co-operative employees for investigation their budgets), VESTNIK STATISTIKI, No 3, p.14-25.

SÄRNDAL, C.-E., SWENSSON, B., WRETMAN, J. (1992): Model Assisted Survey Sampling, Springer series in statistics, New York, Springer Verlag.

WOLTER, K. M., Introduction to Variance Estimation, (1985): Springer Verlag, New York, Berlin, Heidelberg, Tokyo.

STATISTICS IN TRANSITION, June 2002
Vol. 5, No. 4, pp. 643–670

HOUSEHOLD SURVEYS – MAIN TOOLS OF SOCIAL STATISTICS

Filofteia Panduru, Andreea Vasile and Silvia Pisica[1]

ABSTRACT

This paper presents the main characteristics of sampling surveys on households, implemented in the social statistics system in Romania during the transition period to the market economy, focussed on Household Budget Survey (HBS) and Labour Force Survey (LFS). Problems referring to the surveys methodological out-look are presented, with a special accent on sampling (sampling design, size, methods), non-response rates, surveys cost, harmonisation with EU standards.

Key words: household budget survey, labour force survey., survey cost.

1. Introduction

Re-designing of social statistical system in Romania

Re-organisation of the Romanian economy in order to comply with the requests of transition to the market economy imposed the re-designing of the national social statistical system.

Being included in the general process of re-designing of the statistical informational system, according with the principles and mechanisms of market economy, social statistics was oriented to assure the needed information for evaluation of social phenomena, to substantiate the programs in social field and to monitor their efficiency.

The social statistical system includes:

- Population statistics (vital statistics; internal and international migration);
- Statistics of living standards/quality of life (households and families; incomes, expenditures, population consumption; living conditions – housing conditions, households endowment with durable goods; environment etc.);
- Labour force statistics (active population, employment and unemployment; wages, employees; work accidents and professional disease; social protection etc.);

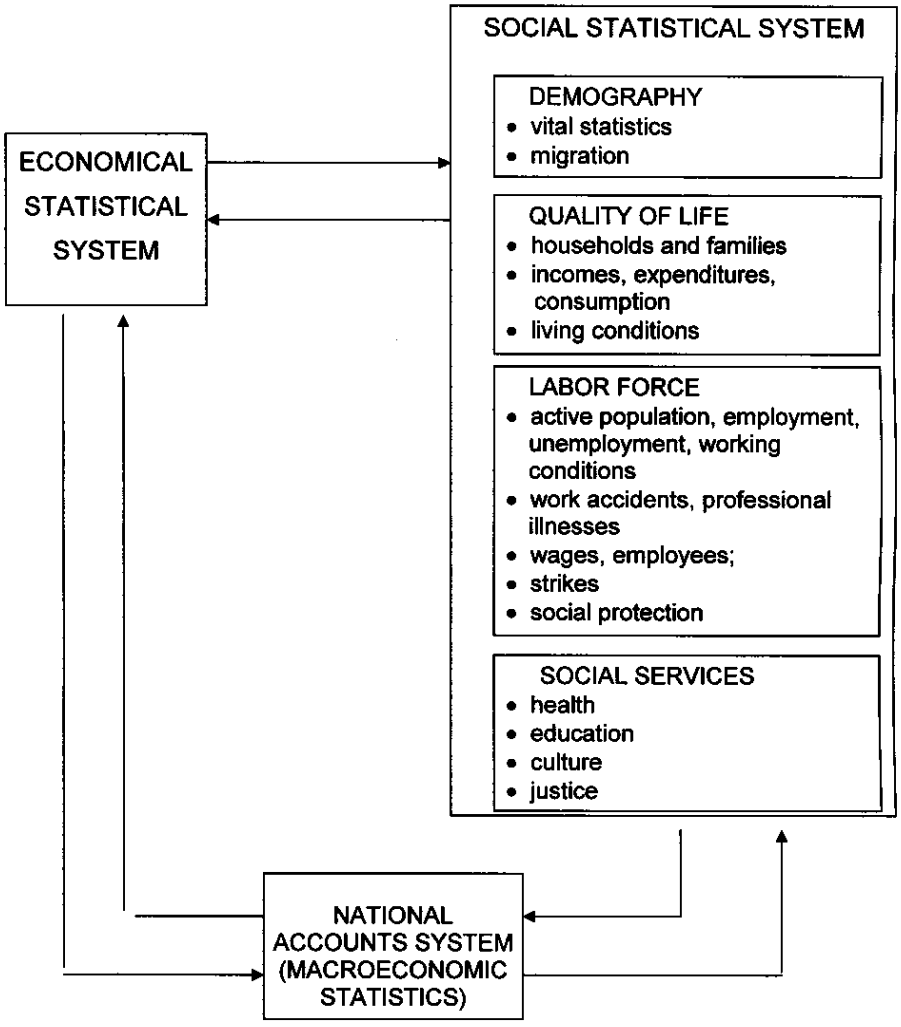
[1] National Commission for Statistics, 16 Libertatii Blvd, 70542 Bucharest 5, Romania.

- Social services statistics (health; education; culture; justice etc.)

An overall overview on the evolution of this system and on the main topics of his re-organisation will allow introducing in the Romanian statistical system some modern methods of investigation, specific for market economy - surveys in households and enterprises.

The major political changes in Romania in December 1989, that implied the transition from a highly centralised economic and social system to the market competition economy, generated normally re-conversion of the statistical system too.

Figure 1. Social statistics in national statistical system



The transition to market economy imposed conceptual and methodological re-organisation of statistical instruments; the most urgent objective was the building of all statistical tools according with the requests and principles of the transition process, but also with the international recommendations and standards. To carryout this objective, it was necessary to take into consideration multiple aspects: specific concrete conditions in our country, availability of human, material and financial resources, options in adopting different solutions, capitalisation of the countries with market economies experience.

Re-organisation of the statistical system took into consideration a well-defined strategy, including a sum of coherent actions, which referred mainly to:

- improvement of indicators system,
- implementation in Romanian statistical practice of methodologies, classifications, nomenclatures, registers and indicators used by the specialised organisations of UN and EU,
- improvement of data collection system through decrease of the number of exhaustive researches and the increase of the selective researches; use of administrative data sources,
- applying of data processing and analyse methods, proceedings of sample's adjustment and weighting of the results,
- statistical investigation of new fields: unemployment, inflation, living conditions, and poverty.

Major changes occurred in social statistics referring to: data sources diversification, data collection and processing methods, adaptation of concepts, definitions and methodologies to the requirements of the transition period and to the European and international standards in this field. Re-designing of the social statistical system supposed firstly a new view of conception of data collection system, changes made in this component being reflected directly by the improvement of the indicators system.

Transition to the market economy necessitated building up the statistical informational system on other premises, radically different from previous ones, on other functioning programmes, other methods and techniques of data collection and processing. In fact, re-organisation of Romanian economy, starting from 1990, represented defiance for the traditional data collection system based mainly on exhaustive statistical researches.

a. Decomposition of big industrial enterprises and appearance of thousands of new units, small and medium size, translated into spectacular increase of the number of enterprises, that made practically impossible an exhaustive periodical research, many of the new enterprises did not have either the discipline nor the expertise needed to fulfil the statistical questionnaires;
b. The deepening of the privatisation process, appearance and the increase of the number of self-employed persons rose supplementary problems, making non-operational the statistical framework of statistical data collection; most

of these workers, not being organised in economic societies or associations, can not be included in researches on enterprises;

c. Appearance of new phenomena (unemployment, inflation) imposed to search new solutions for information collection, but also for their processing, correlation and analyse.

All these considerations and many others, related to the re-adaptation of the indicators content and improvement in using the statistical data, motivated the necessity to re-design the informational system and to **implement surveys in households and in enterprises as main sources of information on social statistics**.

Present social statistics comprise information provided by two data categories:

- exhaustive researches – sources or administrative records and statistical sources such as censuses;
- sampling inquires – statistical sources such as surveys.

♦ **Administrative sources** are used for completing the statistical system with supplementary information necessary to build images as complete and coherent as possible on different phenomena. These sources are specific to many domains of social statistics such as: health, education, culture, and justice. The study of demographic phenomena (nuptiality, live births, infant mortality, and migration) has an old tradition in Romanian statistics and it is based on administrative records also (registers of civil registration offices).

Administrative sources are used as complementary data sources, on the statistics of living standards and on labour force statistics, also.

In social statistics, in general, the administrative sources are used to complete the statistical system with some supplementary information and for checking the estimation obtained through sampling statistical researches. In this category of sources are included: accounting balances and other records of the financial administration, agricultural registers, records of unemployment offices and record of social protection system in general, those of Labour Chambers, documents of civil registration offices.

♦ **Statistical sources** consist in censuses and sampling surveys.

These assure the information collection and processing that cover the main domains of social statistics and allow structuring a system of indicators used to characterise different aspects of social life.

♦ **The censuses of population and dwellings** (PDC) represent the main source of data regarding of population and dwellings. Last dwellings and population census from 1992 assured:

- statistical data needed for describing the changes occurred on the structure of population, households and dwellings, labour resources and labour force,

- a rigorous data base for calculations between censuses on demographic and labour force fields,
- Information needed for construction of a sampling base for designing and carrying –out household's surveys.

Designing and implementing the Multifunctional Master Sample of Territorial Zones (EMZOT), using the results of the 1992 population and dwellings census, represented an important step in re-organisation and re-designing of the national statistical system. In absence of this sampling base it would be impossible to carry out probabilistic surveys and it would not be possible to harmonise the system of surveys in households with the European information collection system.

♦ **The surveys in households** represent important information sources about population living standards, about employment and unemployment.

♦ **System of surveys in enterprises** allows an integrated outlook the main problems of statistics regarding wages and collateral statistics. The system includes current surveys, with a big frequency, in order to fulfil the information needs on short term and surveys with small frequency that supply detailed information for analyses' needs of long term trends.

General background of development of surveys in households

Households surveys are carried out on samples selected from a sampling base built on the 1992 Census results. The sampling base is a Multifunctional Sample of Territorial Zones (EMZOT), that functions on the master sample principle used in other countries also (France, Portugal).

Corresponding to the territorial distribution of EMZOT, a permanent network of interviewers was made up and functions in order to carry out in practice the household's surveys.

Using of this complex sampling system supposes to apply some specific methods of samples adjustment and improvement of estimation quality. All the surveys that are carry out on the EMZOT basis are harmonised from the point of view of sampling design similarity and of common logistic elements used.

EMZOT was built extracting several samples that included 500 up to 2500 census sectors out of total number of 93037 census sectors. The hypothesis that was the starting point in master sampling's carrying out (EMZOT) was the following: this sample, with the correct dimensions and well organised, can highlight the structural and behavioural changes of population in time, with a low cost per observation unit. Thus, within the census's preparing activities, the localities were separated in 93037 sectors (named further Census Sectors – CS), with different dimensions, respectively 100 -110 dwellings in urban areas and 80 - 100 dwellings in rural areas. CS-s were grouped at the localities level in census sections (named further in this paper Census Areas – CA). At one first step, 505

CA-s were extracted using coil method. Out of these, due to their different dimensions, four of them were excluded, those with extreme values. Out of total number of CA-s, these 501 zones represent around 3% of country's population.

Using simple random sampling on first stage assured an implicit stratification of the master sample on the two residence areas (urban and rural). However, this fact was verified on preliminary data as well as on final results from the census.

Through field delimitation 249 rural areas and 252 urban areas were established, each containing 480 and respectively 528 dwellings, with a representation quota of 1 dwelling to 30. Strict delimitation of these areas had in view the proportional allocation of master sample, according to the well-known relation:

$$\frac{n_h}{n} = \frac{N_h}{N} \tag{1}$$

where: n - number of persons from master sample;
N - number of persons of total population;
h - index of the two strata (urban and rural).

It is understood that N is adjusted based on demographic and migration data obtained from statistical and administrative sources. The areas were split into smaller areas of three neighbour dwellings (called on short as zones). Such a splitting, according to which the dwelling to be neighbour at the level of each locality selected is strictly controlled, allows to build samples of different sizes from one survey to another. For these samples randomly with equal probabilities it may be selected either only zones of three dwellings either clusters of two, three or four zones, in this way diminishing the cluster effect. In the same time, this splitting allows to use some rotational patterns according to which, parts of sample to be maintained for several years but over a year not for more than one month.

Periodically, till the next census, the master sample is updated, through re-investigation of those areas and zones limited at zero moment (January first 1992). In this way, the inventory of appearance and/or disappearance of dwellings are made, comparing with the precedent reference moment. From the new dwelling volume, an independent sample is selected for each survey.

The administration system of master sample was conceived in a way that allows monitoring strictly the selections made in time and space, excluding the overlap of the same dwelling in several surveys in the same time or many times and also identifying some characteristics of households in order to treat the non-responses.

The design of one sampling system with multiple uses in Romania was justified by non-existence of some registers that could be used as sampling frame for probabilistic sampling design needed for surveys in households. The

population register in Romania is still in the phase of project and will not be possible to be used for sampling design needed for surveys in households. Other existent data sources are not used because:

- cover only partially the reference statistical population,
- have an heterogeneous content of information referring to the elements of statistical population,
- are not using the computer technique or use only partially it,
- contain registration errors,
- the possibility to update them with reasonable costs doesn't exist.

In the initial form (after re-dimension of centres at the end of 1993) EMZOT included over 257,000 dwellings being used for drawing the samples needed to carry out all the households surveys till 2000.

In 2000, EMZOT was re-built entirely, including 127,000 dwellings for sampling surveys that will be carry out till the results of 2002 census will be available.

Household Surveys

Until 2000, in Romania were carried out two household surveys, which gave statistical information on basic fields of social statistics.

- *Household Integrated Survey* was designed with World Bank technical assistance, through Cornell University experts, USA. The survey was carried out experimental in March 1994 and definitive survey begun in April in the same year. The survey was realised in the period 1994-2000, using waves sampling procedure, each wave containing 3000 dwellings was maintained in the survey one month per year. The randomly selection of the waves was made from EMZOT. The survey functioned using the integration principle of a great number of variables at the level of questionnaire, covering a large variety of information: incomes, expenditures, consumption; endowment with durable goods; living conditions; employment and unemployment; heath; education. The survey objective realisation was based, in the first place, on the linkage between variables at the level of each household. The survey answered to the informational requests on population incomes and consumption, for poverty measurement until the year 2000.
- *Labour Force Survey* was designed and carried out with technical assistance of INSEE experts – France in the period 1991-1993. An experimental survey was organised in October 1993 being operational beginning with 1994.
- Beginning with the year 2000, under Phare National programme RO9703-01: Social Statistics, was made a large project focused on a survey system realisation including the following components:
  - **The updating and enlarging of sample base EMZOT**. The project, realised with technical assistance of INSSE experts, had as main output the realisation of a complementary sampling base and of a new sampling

base with new dwellings which will be used for all households surveys sample selection until the next population and dwellings census.

Household Budget Survey was designed with technical assistance of ISTAT experts – Italy. The survey was realised as pilot survey on a sample representative at the national and regional, 18000 dwellings, in the period May – October 2000. The survey integrates at the questionnaire level information referring to: the size and structure of households; income sources and expenditures destinations; household's consumption by food, non-food products and services. For data on incomes, expenditures and consumption collection is used balance system that allow the data quality checking at the interviewers level. The survey results will be used for population living standard studies, for poverty measurements and social exclusion analysis, and will ensure the elements needed at weighting system on Consumption Price Index calculation. The concepts, definitions, classifications and nomenclatures, and sampling procedures used were harmonised with similar surveys in European Union member's states.

- ⇨ **Living Condition Survey** was designed with technical assistance of ISTAT experts. It was realised as pilot survey in June 2000 on a sample of 10500 households, representative at the national and regional level. Survey provides necessary information for population living condition analysing: living conditions, work conditions, health status, life safety, goods and persons insurance possibilities, endowment with durable goods.
- ⇨ **Health Survey** (technical assistance of ISTAT experts) was realised as pilot survey in June 2000, on a sample of around 9000 households. The survey offers to user information referring to the population health status, medicines consumption, life style, tobacco and beverages consumption, nutrition, chronically diseases and current health problems.
- ⇨ **Time Use Survey** (technical assistance of ISTAT experts) was realised as pilot survey in July-August 2000 on monthly sample of around 4500 households. The survey provides information for analysis of time use modalities by population categories.

Persons involved in survey realisation

National Institute of Statistics has the co-ordinating role and the 42 territorial divisions, proportional dimensioned with the region size, provide a good organisation and supervising of collection, processing and analysis activities.

At the level of each region, the staff involved in survey realisation is structured on three levels:

- Interviewers,
- Supervisors,
- survey responsible.

Supervisors (1 per 5-6 interviewers) are controlling and guiding interviewers activity, are verifying questionnaires fulfilled by enumerators, coding variables to be coded from individual questionnaires and solving the errors occurred during the data entry.

Survey responsible persons (one for each county) co-ordinates entire activity carried out in the county, insuring the control and checking, validation and data transmission to NIS.

At central level, in National Institute of Statistics, small teams of about 2-3 statisticians specialised on the fields covered by each survey, operate the following activities: general control and checking, data validation and analysis on survey results.

2. Household budget survey

History & background

The Household Budget Survey has an old tradition in Romania, the survey begun in the year 1948. The survey begins on 8000 families' sample in urban area. This sample contained families of workers, intellectuals, clerks, self-employed persons, and other type of families covering the whole country territory. The sample of household budget, both in urban and in rural, transformed in the period 1952-1994 (table 1).

From 1952 the survey organisation was enlarged with rural area on a sample of around 3000 peasants households.

Table 1. The sample of households budget evolution in the period 1952 – 1994.

Period	1952-1975	1976-1989	1990-1994
Sample size (families)	7000	11000	9000

In 1994 was designed and realised Household Integrated Survey (HIS) – a multipurpose survey which functioned in the period 1995 – 2000.

In 2000, under National Phare Programme RO-97 was carried out the Household Budget pilot Survey, and beginning with 2001, Household Budget Survey become a permanent survey, which represents the main statistical instrument for population income and expenditures evaluation. Based on concepts and definitions harmonised with European standards, but also with Romanian legislation in this domain, the survey will be able to answer to a lot of acute requests referring to the integrity, coherence and comparability of data.

Survey objectives

The survey offers necessary information for population standard of living evaluation allowing:

- identification of poor households and persons in the purpose of the elaboration of some suitable programs of social protection,
- analysis of the impact of the application of different programs in the social field, of the economical effects of the transition, in general, over the economical and social situation of the population.

The survey provides necessary information for:

- determination of monthly consumption expenditures in COICOP structure, used as weights in CPI calculation,
- final consumption calculation in national accounts,
- identification of social inequalities, of poverty incidence and profile;

Questionnaires

Two types of questionnaires are used for data collection in the Household Budget Survey: Household Questionnaire and Household diary.

The **Household Questionnaire** is structured on 12 sections (modules), which cover all the indicators recommended by Eurostat in the manual "Household Budget Surveys in the EU. Methodology and recommendation for harmonisation". Also, are collected information on health and level of households needs satisfaction compared with their incomes, which are request to be included by different local users.

The content of the household questionnaire is the following:

- **"Household Composition"** which includes questions related with demographic aspects of households' members;
- **"Activity fulfilled during last week"** with questions addressed only to persons from the household, aged 15 years and more, referring to the economic activity in the reference period;
- **"Recipients of pensions, allowances and other social protection benefits"**. The questions from this section offer information on number of persons beneficiary of social protection benefits, number and type of social protection per person beneficiary, the amounts received per each type of social protection benefits;
- **"Agricultural products and foods, alcoholic beverages – balance sheet"** in which are registered stocks, incoming, outgoing in and out of the household, and the household consumption by food and beverages products and group of products (COICOP structure) - quantity and value;
- **"Non-food purchases"** in which are registered purchases of non-food products (quantity and value);

- **"Expenditure for service payments"** where are filled in the amount paid by the households members state, private units and population for services delivered;
- **"Income and other monetary inputs"** where are registered monetary income obtained by household members in the reference month, by sources and also the evaluation of gratuities received by the employed persons from employers;
- **"Expenditures and other monetary outputs - balance"** where are filled in information on households expenditures by groups of expenditures by their destination (consumption expenditures, expenditures for investments, production expenditures etc.);
- **"Other persons than household members who took meals in the household"** containing information related with: number of persons other than household members who ate in the household, number of days of eating in the researched household by type of meal;
- **"Household members who took meals in other household"** containing information related with: number of household members who took meals in other household, number of days of eating in the other household by type of meal;
- **"Meals taken in catering units"** containing information related with: number of household members who took meals in restaurants, cafeteria, snack bar, as guests of researched household, by type of catering unit;
- **"Information on dwelling"** containing questions addressed to the household head and are referring to the characteristics of the main dwelling;
- **"Endowment with durable goods"** is designated to collect information on household durable goods that are still working;
- ***"Health"*** *offer information on household members health, type of illnesses, type of handicap and the expenditures made for the health care;*
- **"Satisfaction of households needs compared with their income"** offer qualitative information on financial situation of the household.

The diary contains the following chapters:

- "Incomes and other monetary inputs of the household members";
- "Expenditures and other payments of the household members";
- *"Incoming agricultural products and foods from various sources"*
- "Outgoing agricultural products and foods by use"
- "Comments of the household".

Survey organisation

The survey is realised on a sample of households from the urban and rural areas, selected randomly from all the counties of the country and from the Bucharest City.

In the research households from all social and economic categories are included, namely: employees, employers, own-account worker in agriculture activity (peasants), and members of agricultural associations, own-account workers in non-agricultural activities (crafts, worker in trade, etc.), members of non-agricultural co-operatives, unemployed persons, pensioners and others.

The object of the registration are all the persons with permanent domicile in Romania, members of the selected households, who are present, temporary absent or who are left for a longer period and who participate totally or partially at the budget of the household (incomes and/or expenditures).

Not included in the survey are institutionalised persons (old persons, handicapped persons asylums, homes for workers, sanatoriums etc.).

The registration of the data in the household's questionnaire was realised on basis of **interview**, and in the household's diary, by **self-registration**.

The filling in the household's questionnaire is made at the domicile of the household, through discussions with each member of the household, and if this isn't possible in all cases, the information were obtained on basis of the declaration of the head of the household.

The filling in the diary of the household is made by self-registration by the head of the household or another adult member of the household who can do correct registrations daily.

The period for which is registered the data in the questionnaires of the survey is the **month** (from the first to the last day of the month). Each month is collected data for 3006 dwellings. The interviewers are spread all over the country and each interviewer has 6 dwellings per month. The collection of the information from the households is done through 3 compulsory visits at the domicile of the households and, if it is the case, through supplementary visits**.**

Survey organisation as Household Budget Survey involves a great consumption of human and financial resources. In the year 2000, for the organisation of the pilot survey on a sample of only 18000 dwellings, was spent around 235000 Euro. More than 50% of the entire amount allocated for the pilot survey was used for field operations (payment of interviewers, of county surveyors, of data entry operators etc.). Also the printing of all the questionnaires and of the two publications use around 18% of the total amount.

In the respect of non-response rate, this was constant in time very low (table 2). In the last years, when was carried out Household Integrated Survey and also for pilot HBS, the response rate began to diminish. Households are given a small amount of money (as around 1,5 Euro) to stimulate them to fill in correctly the diary for one month.

Table 2. The evaluation of response rate for household survey in the period 1998-2000.

	Accept %	Refusals %	Not at home %	Not found dwelling %	Not occupied dwelling %
1998	89,56	4,57	3,30	0,27	2,30
Urban area	84,63	7,49	5,29	0,33	2,25
Rural area	94,70	1,53	1,22	0,21	2,35
1999	88,37	5,45	3,29	0,26	2,63
Urban area	82,44	9,16	5,49	0,36	2,55
Rural area	94,52	1,60	1,01	0,16	2,71
2000	87,52	5,98	3,25	0,38	2,87
Urban area	80,90	10,22	5,62	0,54	2,72
Rural area	94,38	1,58	0,81	0,21	3,02

Sample design

The Household Budget Survey (HBS) is organised as a continuos survey each year, on a sample of 36072 dwellings, distributed in monthly independent sub samples of 3006 dwellings. The sample is representative at national and regional level (development regions). Sampling plan has two stages.

Sample selection

On first stage "EMZOT master sample" is selected, realised on the basis of the data registered at the Census of the Population and Dwellings from 1992. EMZOT is a very big sample (about 250000 dwellings) and was realised with the purpose to be used as sample base for household surveys between two censuses. It is a systematic sample of 501 geographical areas, called research centres, distributed in all the counties, both in urban and rural area. . These 501 research centres are assimilated with *sampling primary units*. EMZOT includes 259 research centres in urban area and 242 research centres in rural area. The representativity of the localities it was ensured using the main characteristics of the dwellings and households: average number of households in a dwelling, average number of persons per household, occupational status of the household head.

On second stage it is selected dwellings from each research centre. Individual dwellings aren't selected directly. Groups of 3 dwellings were selected based on a randomly selection algorithm. The name of these groups is clusters and is assimilated with *secondary selection units*. The algorithm of randomly selection

was based on calculation of selection interval and a random start for each research centre. From each research centre are included, monthly, in the sample two clusters, 6 dwellings (24 clusters, 72 dwellings for 12 months). For the sample dimensioning in the second stage 36072 dwellings was taking into account to obtain estimations for the main characteristics of the survey which could be affected by errors in the limit of 3% and guaranteed with a probability of 97%.

Weighting method

The results at the country level are grossed up based on the coefficient assigned to the persons from the household from sample, who have answered to the interview.

When setting up these coefficients, the following steps were taken:

a) Calculation of the basic weights;
b) Non-response adjustment;
c) Final adjustment of the sample population and calculation of the final weights.

a. Calculation of the basic weights

The basic weights of the dwellings (households) represent the first estimation of the grossing up coefficients. For their determination it was needed the following steps:

- The calculation of the sampling primary units selection probabilities from EMZOT, different for urban and rural areas *($P_{1U,}P_{1R}$).*

For urban area:

$$P_{1U} = \frac{n_{1U}}{N_{1U}}, \tag{2}$$

where: n_{1U} - number of dwellings from urban area included in EMZOT;

N_{1U} - number of total dwellings from urban area;

For rural area:

$$P_{1R} = \frac{n_{1R}}{N_{1R}}, \tag{3}$$

where:n_{1R} - number of dwellings from rural area included in EMZOT;

N_{1R} - number of total dwellings from rural area;

- The calculation of the sampling secondary units selection probabilities from the each research centre *k* ***(P_{2k}).***
  These probabilities are, in fact, the inverse of the sampling interval for a research centre.

$$P_{2k} = \frac{1}{I_k} = \frac{n_{2k}}{N_{2k}}, \tag{4}$$

where: I_k - sampling interval for a research centre k;
n_{2k} - number of dwellings in sample from the research centre k;
N_{2k} - number of dwellings in the research centre k.
$k = 1,\ldots.,501$

- The calculation of the selection general probabilities of the dwellings in the research centre k, on urban and rural areas (PGS_{kU}, PGS_{kR}).

For urban area:

$$PGS_{kU} = P_{1U} \; x \; P_{2k} \tag{5}$$

For rural area:

$$PGS_{kR} = P_{1R} \; x \; P_{2k} \tag{6}$$

- The calculation of the dwelling (household) basic weights in the research centre k, on urban and rural areas (PB_{kU}, PB_{kR}) .

For urban area:

$$PB_{kU} = \frac{1}{PGS_{kU}} \tag{7}$$

For rural area:

$$PB_{kR} = \frac{1}{PGS_{kR}} \tag{8}$$

The basic weight (PB_k) is the first estimation of the extension coefficient. This basic weight is a weight for all dwellings in the research centre k, but it is used also to the households living in these dwellings and implicitly to all the persons from the households.

b. The non-response adjustment

To cover the percentage of the households that refused to participate at the survey, it was adjusted the basic weights of the households obtained in the step before. For this it was done:

- Homogenous cells of households from the respondents sample, in which it was considered that the probability of response is the same for all individuals. These homogenous cells were made using the combination of the following variables: region, county, area and occupational status of the household head.

- Adjustment rate with non- response for each cell. In this way, the households that were grouped using these cells and the non-responses were treated separately on each cell.

$$RNR = \frac{\sum PB\ ES}{\sum PB\ RASP\ ES}, \tag{9}$$

where: RNR – adjustment rate with non-response;
∑ PB ES - the sum of the basic weights of the households from the sample;
∑ PB RASP ES - the sum of the basic weights of the households from the sample, which answered at the interview.

The calculation of the **adjusted weights with non-response rate** assigned to each household and each person from the same household was as follows:

$$P_{AJ} = PB \cdot RNR \tag{10}$$

c. Final adjustment of the sample population and calculation of the extension coefficients

The final adjustment was made to ameliorate the estimations, focusing at the adjustment of the sample in the case, when appear differences between the respondents sample structure on some interest variables and the total population structure on the same variables. For this auxiliary information, as demographic variables (sex, groups of age) and localisation variables (area of residence, region) disposable from demographic sources (legal population at first of July) was used. The final adjustment was made using CALMAR software, on the following levels:

Regions, areas of residence, sex, groups of age: 0-14 years, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75 and over.

The final adjustment of the sample is based on the algorithm ranking-ratio. For each region cells of persons crossing the variables are made: area/sex/groups of age, from the sample and from the total population. For each region, in each cell area/sex/group of age is multiplying the weight obtained in the steps before, with a coefficient of final adjustment (C_r):

$$C_r = \frac{\text{Total number of persons by region / area / sex / group of age}}{\text{Number of weighted persons from sample on region / area / sex / group of age}} \tag{11}$$

At the end of this step, the final adjustment coefficient assigned to each person from household is:

$$COEF_{ext} = P_{AJ} \cdot C_r = PB \cdot RNR \cdot C_r \tag{12}$$

Data processing

The data processing and validation is made in two steps:

- decentralised processing – at the county level which include manual codification of the questionnaires, computer data entry , logical checking and data validation in the research centres;
- centralised processing –validation of the data from the county level, centralised processing and elaboration of tables with results for the total country.

Programmes for data entry and logical checking of information from questionnaires find errors and edit list of errors to be analysed and corrected.

The error corrections are made taking into account the type of error and the motive of apparition. The errors could appear because of:

- mistake of data entry operators – in the case of wrong typing of data from questionnaires;
- wrong codification of information from questionnaires.

After launching the programmes for logical control checking, two types of errors are included in the error list:

- refusal – logical conditions that are not accepted by the statistician;
- attention – in the case when the problem could exist in the reality as exception and could be accepted by the statistician after analysing the problem.

Dissemination of the results

The diversity of information collected by Household Budget Survey allows the possibility to make profound analysis on the economic and social situation of household or members of the households. The main publications where data from HBS are contained are:

- detailed publication containing a short methodological part, analysis of data and tables annexed;
- Statistical Yearbook;
- publications on poverty;
- CESTAT bulletin;
- Social Trends.

In fact, data from this survey in all publications contain information on income and expenditure.

3. Labour force survey

History & background

In October 1993, a pilot survey was performed on a reduced sample of 10 000 dwellings comprising 40 selected clusters, 19 in rural area and 21 in urban area. The aim of pilot survey was to check and evaluate the questionnaire and the

fieldwork, as well as the concordance between age group distribution of the sample population and distribution of total population as it was registered during the census from 1992. In 1994 and 1995, two annual Labour Force Survey (LFS) were carried out in March for two reference weeks. In 1994 the sample size was 15030 dwellings and in 1995 18036 dwellings. Main reason for increasing the sample size was the criterion of improving the estimations for certain categories of population, which, during the pilot survey did not have a significant proportion.

In January 1996, a continuous LFS started, providing quarterly and annually (averages) results. The quarterly sample size is about 18036 dwellings. A two-stage probability cluster sample of housing units is used. In the second stage a 2-2-2 rotation pattern is used. The rotation pattern was chosen in order to accomplish the following two main criteria:

- reduction of surveyed household burden, respectively of non-response;
- creation of a household panel with the aim of studying the occupational mobility of population and the average duration of unemployment.

In 2001 survey questionnaire, concepts and definitions were redesigned and revised. Thus, starting from 2002, new variables were introduced, such as: atypical work, homework, supplementary information on forms of training and variables corresponding to the new operational definition of unemployment according to Eurostat recommendations. The quarterly survey was redesigned in order to follow the international concepts and definitions, Romanian legislation in force and to reach the harmonisation, in terms of concepts and methodology, with the market economy countries.

Survey objectives

Household Labour Force Survey is a modern procedure for the statistical inquiry of labour market, having as main objective the measurement of active population - employment and unemployment – and of inactive population.

Conceived as an important source of information on labour market, this survey provides, in a consistent manner, essential data on all population segments, with a lot of possibilities of correlation and of structuring according to various demographic, social and economic characteristics, under international comparability circumstances.

Quarterly continuous since 1996, LFS allows the achievement of conjuncture data on the size and on the structure of labour force supply, pointing out the seasonal phenomena manifested on labour market.

The advantages of this statistical source, listed below, are mainly entailed by a situation where the observation unit is the household:

- it offers the possibility of people observing, not only from the standpoint of their appurtenance to labour force, but also as members of a household;
- provides information on certain categories of population which could not be obtained from other sources, such as: under-employed persons and

discouraged people;
- does not depend on administrative sources, therefore is not influenced changes of administrative and legal nature;
- the survey methodology is adapted to Eurostat recommendations, thus being possible the comparability with other EU countries.

Questionnaire

Data are collected with identical questionnaires through the whole year. The questionnaire is divided into two parts:
- dwelling questionnaire;
- individual questionnaire.

The first consists of:
- a part concerning information about the building and the dwelling: its destination, situation (previously surveyed, new in the sample, has to be excluded from the survey: destroyed, changed the destination, disappeared by fusion etc.), type (permanent, temporary), number of households living in;
- a part for each household, addressed to all household members disregarding the age, collecting demographic and educational data, information concerning the relationship with the head of household, mobility in comparison with the previous survey and the presence in the household.

The individual questionnaire, addressed only to the household members aged 15 and over, consists in several parts:
- general questions addressed to all persons aged 15 years and more (15 years was adopted as minimum age limit for defining the active population) which allow to guide the interview to the next modules of the questionnaire, depending on economic activity of the respondents;
- questions addressed to persons who worked regarding their professional status, place of work, occupation, main and secondary activity, hours worked, homework and atypical work, willingness of another job);
- questions addressed to persons who did not work (unemployed and inactive) regarding their professional status, activity, occupation and place of work before stop working, willingness for a job, reasons for not seeking job etc.;
- questions addressed to persons who are seeking job (unemployed or employed who are seeking for another job), methods used for seeking job, availability for work etc.;
- questions regarding education and training level, type and other characteristics of training during last four weeks;
- questions regarding geographical and social mobility.

Survey organisation

There is no specific national legislation concerning obligation to provide information for the survey, and participation in the survey is voluntary.

The whole country is covered and only the private households are surveyed.

The target population includes all the Romanian citizens, residents in Romania. The survey covers all the members of the selected households including the persons absent from home for a longer period (over 6 months), if they are preserving family relations with the household to which they belong, such as:

- military on duty (conscripts),
- pupils and students away for study,
- persons left for work,
- prisoners,
- persons temporary in hospitals or sanatorium for medical cure.

The data are collected only by face-to-face interviews - registration on paper-and-pencil. Main role in surveying belongs to the 501 interviewers who have to visit each household from selected dwellings and to record the answers of each person member of the occupied household.

Questions and recorded answers are referring to the situation of the interviewed persons for a certain period of time, i.e. reference period.

Main reference period is the **week**, from Monday till Sunday, which precedes the registration week. But some topics, such as atypical work, homework, education or training received have four weeks as reference period; main labour status has three months as reference period etc.

There are 52 weeks as reference period. Sample was evenly spread throughout entire year in order to ensure a continuous data collection.

Overall non-response rate is quite low, even if it increased from 6.4% in 1996 to 8.9% in 2000. Non-response rate due to interview refusals is much lower, between 0.9% in 1997 and 1.5% in 2000. When the identified (*targeted*) person is not available to answer, the information is collected by *a proxy* (i.e. someone else – in the household or not – who answered the questions on behalf of the *targeted* person). Percentage of proxy interviews is about 25% – 28%.

Total cost of the survey is about 255000 Euro. Out of total, around 40% represent field operation costs. More than a quarter (28%) is the expenditures for printing the survey tools (questionnaires, letters containing information to respondents sent in advance, interviewers' handbooks, quarterly and annual publications on survey results etc.). For LFS no incentives are paid.

Sample design & rotation pattern

The survey is carried out based on a randomly selected sample of dwellings, representative of country level. The quarterly sample size is about 18036 dwellings (Ultimate Sampling Units – USUs), with all component households. A two-stage probability cluster sample of housing units is used.

In the first stage, a sample of 501 areas, Primary Sampling Units (PSUs), was designed after the 1992 census. This is the Multifunctional Sample of Territorial Areas, so called the master sample EMZOT, selected from each county (42 counties) with probability proportional to size, given by population (number of residents). The EMZOT sample has 259 PSUs from urban areas and 242 PSUs from rural areas.

In the second stage clusters composed of three housing units each, were systematically selected.

A 2-2-2 rotation pattern is used. One dwelling is observed for two consecutive quarters, then temporarily removed for the next two quarters and enters again for the next two quarters afterwards being definitely removed from the survey. Thus, the households of each selected dwelling are interviewed for 4 times. According to the procedure in use, during each quarter are kept in the sample 50% of the dwellings observed in the previous quarter, 25% of those observed two quarters ago and the remaining 25% are new dwellings. Thus is also preserved 50% coverage of a quarter sample as against the same quarter of previous year.

Weighting procedure

There are three main steps to the estimation process:

a. basic weighting;
b. non-interview adjustment;
c. raking ratio adjustment (Iterative Proportional Fitting).

Formulas applied are the same with those presented at previous chapter (HBS).

a. Basic weighting:

The basic weighting procedure begins the process of weighting or inflating the sample data to produce an estimate for the entire population. Each housing unit (HU) has a probability of selection. The inverse is the base or basic weight. All persons in a sampled HU have the same basic weight.

b. Non-interview adjustment

Each quarter, about 4%-5% of occupied housing units is not interviewed due to unavailability of respondents, refusals etc. In the non-interview adjustment procedure, the weights of all interviewed HUs are adjusted to account for occupied households not interviewed.

Data from interviewed persons in each cell are weighted up to also represent sample persons in the same cell who were not interviewed.

At the end of non-interview adjustment procedure, the weight associated with each HU and person is the product of base weight and non-interview adjustment.

The weights and adjustments are used to inflate sample data to represent the entire population.

c. Ratio adjustment/Linear adjustment

Distribution of county, area of residence, sex and age in the sample differ from the distributions of these characteristics in the total population.

Biannually, independent figures of population distributions are produced based on demographic sources. Because these characteristics are closely correlated with labour force participation and other principal measurements made from the sample, the survey estimates are substantially improved when weighted appropriately by the known distributions of these population characteristics.

The ratio adjustment procedure reduces the variability of estimates from the sample. Ratio adjustment results in weights for sample person records that produce sample estimates consistent with some population controls, the known distributions of the population characteristics, so called "independent cell estimate":

County (42 counties) / **Residence** (urban, rural) / **Sex** (male, female) / **Age** (15 groups).

Ratio adjustment proceeds are as follows: for each county/residence/sex/age cell multiplies the person weights through non-interview adjustment by the following ratio.

$$\frac{\text{Independent county/residence/sex/age cell estimate}}{\text{Weighted county/residence/sex/age estimate from LFS sample}} \quad (13)$$

The ratio adjustment procedure force LFS county/residence/sex/age cell estimates to equal the independent cell estimates. This procedure is repeated for a maximum of fifteen times. Each time the step above is executed, the computed ratio gets closer and closer to 1.0. After maximum fifteen iterations the ratios are nearly all equal to 1.0, insuring virtual equity between LFS sample estimates of the population and the set of independent estimates.

At the end of the ratio adjustment procedure, person weight is the product of: base weight, non-interview adjustment and ratio adjustment.

Additional topics to the main survey

For a better characterisation and for deepening the phenomena manifested on labour market, several complementary surveys (ad hoc modules) were planned to be carry out during a single quarter every year, on different topics:

- 2000: Transition from school to working life;
- 2001: Length and patterns of working time.

Ad hoc module on ***Transition from school to working life*** was implemented and carried out during third quarter 2000, having as main objective to provide the information needed for the characterisation of job seeking after leaving the

education and youth integration on labour market. Persons aged 15-35 years old who left education for the first time during last 10 years were the target population of the complementary survey. In 2004, the ad hoc module on this topic will be repeated.

Ad hoc module on ***Length and patterns of working time*** was carried out in second quarter 2001. Main objective of this survey was to provide information on length and characteristics of working time arrangements. A subject of interest was referring to the proportion of those who are working according to so called standard programme of work (5 days per week, 8 hours per day), comparing to the proportion of those who are working according to non-standard arrangements (variables hours, working time banking, on-call, min-max contract hours etc.). Population targeted for this survey was all persons aged 15 years and over that had as professional status on main activity:

- employee;
- employer (self-employed with employees);
- self-employed (without employees), including members of co-operatives.

Within the International Programme for Elimination of Child Labour, over one year (October 2000 – September 2001), a complementary survey was carried out, attached to LFS.

Target population of this survey was children aged 5-17 years old. This survey was addressed to one adult – mother, father or another persons who takes care of the child – and to each child from the age group already mentioned.

The ***modular child labour survey*** was envisaged to go beyond the statistical count of the number of economically active children. Survey data allows the study of the scale, distribution, characteristics, causes and consequences of child labour. It provides needed information on children engaged in non-economic activities (school attainment, housekeeping activities) and comprehensive demographic and socio-economic characteristics of all school-age children, and (for working children) working conditions, safety and health aspects (focussing on type, frequency, and gravity of injuries/illnesses), and reasons for working. The survey also identifies the demographic and socio-economic characteristics of parents of any child in the 5-17 age bracket, since there is correlation between these and the incidence of child labour.

Data processing

IT solution for the statistical survey is accomplished using Visual Fox and is divided on two components:

- IT solution at local level
- IT solution at central level.

To design the IT solution at local level 3 categories of logic control were established:

- checking of logical ways through the questionnaire;
- checking of questionnaire data;

- checking the integrity between the identification data of selected households and of those for which the data were filled in during the interview.

Control conditions for each questionnaire requested respecting the flows/logical ways through questionnaires, correlation between answers to different questions, erasing illogical and aberrant answers.

Two procedures were created: logic control of data files and processing data to obtain the tables.

IT solution at local level allowed data entry from fulfilled questionnaires and it validation at regional level.

IT solution at central level contains the alternative for regional level and the IT programs for:

- preparing the local databases for database aggregation;
- automatic corrections;
- printing the alternative "another" for answers;
- straightening out and extension;
- editing the tables results;
- automate codification of national variables in accordance to Eurostat list;
- creating the files for automate loading of data in statistical database.

Dissemination of survey results

Main publications on survey results are:

- a quarterly detailed publication that contains survey methodology and organisation, results analysis and tables; published in Romanian and English;
- annual detailed publication that presents survey methodology and organisation; it contains analysis and average annual data, as well as information on the evolution of main indicators during period of 1996 – 2000; published only in Romanian;
- Statistical Yearbook;
- CESTAT Bulletin;
- Social Trends;
- Human Development Report.

Survey results are available on electronic format also.

4. Options for 2002-2007 Time Horizon

New system of Romanian social statistics has two major components:

♦ **Data sources with total enumeration character** composed from: **specific inquires of public statistics** (census of population and dwellings, statistics on marital status for vital statistics and population migration, statistics on education, health, culture) and **administrative data sources** which has to be more

integrated in the public statistics system. In this respect, the essential problems that has to be solved are referring to the following main aspects:

- Adopting the definitions, classifications and unique nomenclatures, harmonised with the ones used at European level;
- Using modern methods of data collection, transmission, processing, analysis and dissemination of information;
- Identifying available data from exhaustive sources referring to auxiliary variables that may be used for improvement of quality of estimations obtained through the sampling statistical surveys.

Population and dwelling census (PDC), planned to be carrying out in March 2002, has to represent an absolute priority of social statistics. 2002 PDC will provide basic information for future inter-census evaluations referring to population and labour resources as well as for creation of a new sampling frame necessary for household surveys which will be conducted till next census.

♦ **System of household surveys** will include starting from 2001:

- **Quarterly Labour Force Survey** conducted as a continuous research, updated and completed in respect of research programme according to the recommendations formulated during the Eurostat working groups. With a certain periodicity, ad hoc modules will be carried out as complementary surveys harmonised with the ones conducted by the EU Member States:
  – Particularities of employment of disabled persons (2002);
  – Lifelong learning (2003).
- **Household Budget Survey** will be conducted as a continuous survey on a waves sample of about 3000 households per month, with a rotational pattern as the level of each year. Together with living conditions survey, it will provide the statistics needed for the evaluation and for monitoring the poverty and social exclusion.
- **Living Conditions Survey**, which became a permanent tool of social statistics, might be transformed over future years in a harmonised survey of EU-SILC type on incomes and social exclusion. Organisational and methodological concept of this survey will be adapted to the requirements of the Regulation that on next future (2002) will be approved by the Statistical Program Committee (SPC) and by European Commission. The Regulation refers to the ways of designing and conducting a harmonised EU-SILC – Incomes and living conditions statistics within EU.
- **Other specialised surveys** will be design and conducted depending on requests and without a planned periodicity. Time use survey, female fertility, population health and others are to be taken into account in this respect. Complementary modules recommended by Eurostat with a given program and periodicity established at EU level will be also carried out.

Priorities:

- Identification of adequate methods and techniques for sample size dimensioning and for determining the estimations on small fields, consequent request of needs for information at county level;
- Designing a new master sample (EMZOT) based on 2002 PDC results;
- Using efficient methods of quality checking of collected data, preparing quality reports;
- Applying mathematical and statistical methods in results analysis and for estimation errors computation;
- Finding new solutions for keeping "human values" under the conditions in which the remuneration system of young statisticians is inadequate and the time and investment in their training is highly significant.

On present structure, with it components and ways of functioning above presented, Romanian social statistics assure organisational, conceptual and methodological frame, providing the information needed and requested by internal and international users in terms and conditions of comparability and coherence. Within general process of re-conversion and restructuring of national statistical system, social statistics were redesigned and developed according to European Union standards and to international bodies specialised in this field. The harmonisation to European statistical system has been done respecting and following main topics:

- research methods;
- sampling plan;
- system of classifications and nomenclatures.

Further development of social statistics will be focussed on fulfilling the request of statistics users, alignment to international recommendations and standards in this field and assurance of obligations assumed by Romanian statistics through the agreements signed with European specialised bodies.

REFERENCES

ARDILLY, P. (1992), Introduction aux techniques des sondages, version 1, INSSE, Paris.

BOGDAN, V. (1991), Sisteme de eşantionaj cu utilizări multiple, Revista română de statistică, nr.1/1991 (Sampling Systems with multiple utilisation, Romanian Statistical Review nr.1/1991).

CHERNYSHEV, I. (1991), Comparable Employment and Unemployment Estimates: a Technical Guide, ILO, Geneva.

DEVILLE, J.C., ROTH, N. (1993), La precision des enquetes sur l'emploi, INSSE, Paris.

EUROSTAT (2000), La statistique sociale europeene redouble de vitesse, SIGMA, Le bulletin de la statistique europeenne, Comission Europeenne, , nr.3/2000.

EUROSTAT (1997), Household Budget Survey in the EU; Methodology and recommendations for harmonisation, EUROSTAT, 1997.

EUROSTAT (1998), Household Labour Force Survey; Methods and definitions, EUROSTAT, 1998.

EUROSTAT (2000), Labour Force Survey in central and east European countries; Methods and definitions, European Commission 2000.

HUSSMANS, R., MEHRAN, F.,VERMA, V. (1990), Surveys of Economically Active Population, Employment, Unemployment and Underemployment; An ILO Manual on Concepts and Methods, ILO, Geneva.

INSTITUTUL NAȚIONAL DE STATISTICĂ (2000), Aspecte privind Calitatea Vieții Populației. Institutul National de Statistică, Bucureşti, 1995-2000 (Aspects on Population Quality of Life, Bucharest, 1995-2000).

INSTITUTUL NAȚIONAL DE STATISTICĂ (2000), Ancheta asupra Forței de Muncă în Gospodării, Bucureşti, 1995-2000 (Household Labour Force Survey, National Institute of Statistics, Bucharest, 1995-2000).

KELETI, A.(1990), Les statistiques de la main d'oeuvre en transition, CES/AC.51/35, Troisieme Reunion Mixte CEE/OIT sur les statistiques de la main d'oeuvre, Geneva, 1991.

OLTEANU, I., BOGDAN, V.(1993), Elemente de teorie şi practică statistică, suport de curs pentru uz didactic (Theory elements and statistical practice, didactic paper), Braşov, 1.

PANDURU, F., GHERGUȚ, D.(1994), Methodological and Conceptual Labour Approaches of the Romanian Labour Force Survey, *Statistics in Transition*, Volume I, Number 4, June.

PANDURU, F., GHERGUȚ, D., GRAMA, Gr.(1994), The Organisation and Implementation of the Romanian Labour Force Survey in Countries of Central and Easter Europe, Budapesta.

PANDURU, F. (2001), Statistica Socială în contextul integrării europene, Seminar ştiințific, "Bună dimineața Europa! Integrarea României în UE", Iaşi, 10-11 mai 2001 (Social Statistics in the Context of European

Integration, Scientific seminar, "Good morning Europe! Romanian Integration in EU", Iaşi, 10-11 May 2001).

PANDURU, F., BIȚĂ, V. (coord.), MOLNAR, M., VASILE, A. (2001), and other, Co-ordinates of living standard in Romania; Population incomes and consumption, National Institute of Statistics, ROMASDAR – Consultants, Bucharest.

STATISTICS IN TRANSITION, June 2002
Vol. 5, No. 4, pp. 671–685

HOUSEHOLD SAMPLE SURVEYS IN SLOVENIA

Vasja Vehovar[1], Metka Zaletel, Tatjana Novak, Marta Arnež, Katja Rutar[2]

ABSTRACT

The paper describes the sample design and implementation of two major household surveys conducted by the Statistical Office of the Republic of Slovenia: Labour Force Survey and Household Budget Survey. Both surveys were redesigned in 1997. The new sample designs have basically followed the Eurostat guidelines, however, they incorporate certain country-specific features, too. The paper also provides the background information about 10-year transition process of the Slovenian Statistical Office from a regional (and "socialist") statistical bureau – performing almost no sample surveys - to a modern European statistical agency. Further on, the sample designs are described, together with the discussion of open issues. Response rates, costs, sampling and non-sampling errors are discussed as well as the specifics of the corresponding survey modes and patterns of panel rotation.

1. Introduction

Slovenia gained independence in the beginning of the nineties, which brought many changes into the public sector, including the role of the official statistics. Before that, within former Yugoslavia, the majority of the official statistical activities were centralized in the Federal Statistical Office. This was particularly true for the methodological work concentrated at the Federal Statistical Institute. This institute was a very professional and competent in contemporary statistical methods. For example, they designed a regular Household Expenditure Survey with sophisticated regression estimator already in the mid-sixties.

In addition to the centralization of projects and methodology, the Federal Statistical Office often performed the data analysis, too. So, the raw data would

[1] Faculty of Social Sciences, University of Ljubljana, Slovenia.
[2] Statistical Office of the Republic of Slovenia.

travel to Belgrade for processing and then returned as output listings. Another problem was that many methods were not implemented in a full range, at least in the range comparable to the developed countries. This was particularly true for household sample surveys. They were used only in a very limited scope. On the other hand, the statistical system was providing a wide range of quality statistical data.

We should also mention another Slovenian specifics – a rich experience with attitudinal sample surveys. The Faculty of Social Science of the University of Ljubljana, for example, conducted the General Social Survey (GSS) on a regular basis since 1965. There exists a comparable series of GSS surveys with almost unchanged sample design for almost 40 years.

With independence the Slovenian Statistical Office gradually transformed from a regional statistical bureau to a modern European statistical office. The process was relatively soft without many radical changes. In part, this was due to unchanged top management, which remained on power for the whole period of more than 15 years. The new features, tools and projects were introduced intensively and systematically, but yet gradually. The new professional staff was recruited and systematically educated. In large part the training and education was performed abroad. Another advantage was a successful cooperation with researchers and professionals from the universities. The Labour Force Survey is perhaps the most typical example of this cooperation: it started at the University of Ljubljana and it was moved to Statistical office only after five years of its initial introduction.

2. The Labour Force Survey (LFS)

1.1. The history of LFS in Slovenia

The first LFS started in 1989 as a research project at the Institute of Social Science, Faculty of Social Science, University of Ljubljana (Vehovar, 1997). In the 1989 this survey included 4,000 persons, aged 15 to 70 years. Persons were selected in a two-stage cluster sample from the population register. In 1990 the survey was repeated with all responding persons from 1989 and with additional 4,000 persons included from a fresh sample.

Following the ILO (Hausman et al. 1990) and the Eurostat working guidelines, in 1991 the Slovenian LFS switched to the survey of households. However, due to legal restrictions, the population register was not available that year. Instead, a three-stage area sample was used: 150 communities – 300 enumeration areas – 3,000 dwellings. All households in the selected dwellings were asked to participate. Compared to the sampling from the register of population the area frame gave similar results, but the costs and complications were dramatically higher (Vehovar, Zaletel, 1995). At the same time the field substitution procedure for nonresponding units (used in LFS 1999-2000) was

abandoned and replaced with weighting adjustments. It was shown in a profound study (Vehovar, 1999) that very rarely there are any gains in filed substitutions. On the other hand, severe disadvantages inevitably accompany this survey practice.

From 1992 to 1996 the sample was composed of responding households from previous two years together with newly selected households. The households stayed in the panel for three years and the sample size stabilized at around 8,500 households.

In 1993 the responsibility for this survey moved from the Institute of Social Science to the national Employment Office. It was also included into the National program of statistical research. In 1993 the register of population again was used as a sampling frame.

In 1995 the Statistical Office of the Republic of Slovenia finally took the responsibility for the LFS survey. In 1995 and 1996 the design remained unchanged as an annual panel with face-to-face survey, which was carried out in May each year. The data collection lasted four weeks with the selected households remaining in a sample for three consecutive years. The reference period for working activities were the last seven days.

1.2. The redesign of LFS sample in 1997

The sample for LFS 1989-1996 was designed and conducted in a rather ad-hoc manner, depending heavily on an uncertain yearly budget. For this reason, with fixed budget the sample was optimizing only the samples size alone and not the precision of the estimates. The priority was to avoid empty cells in tabulations as opposed to obtain optimal precision and reasonable design effects. Due to only 150 primary sampling units (PSUs) the design effect for the unemployment rate was extremely high, around deff = 3.

a) Reasons for redesign

In 1997 a major redesign took place. Main reasons for the redesign were as follows:

- the need for more frequent results (e.g. quarterly instead of annually);
- the increased need to harmonize with Eurostat requirements;
- the need for more detailed results, e.g. users need data not only for the whole Slovenia but also for the regions;
- the request for improved precision (design effect) of the estimates.

According to a new sample design the LFS became a continuous panel survey with quarterly sample selection as well as with quarterly publication of results. Each quarterly sample was divided into six two-week intervals. Reference period for the interviews was the week (from Monday to Sunday) before interviewing. An important aspect of the redesign was also the comparability with the results from previous year.

b) Rotation model of the panel

The panel design provides data for longitudinal analysis – the transition tables. In a yearly survey (until 1996) the overlap between two consecutive years was 60 %. With a quarterly survey the intention was to keep the high overlap between consecutive years and to establish it also for the consecutive quarters. At the same time, the household interview burden should be minimized.

For this purpose the rotation model 3-1-2 was applied. The households were interviewed three consecutive quarters, than omitted for one quarter and included again for another two quarters. This model was applied also at Statistics Finland.

Table 1. LFS rotation model (3-1-2) in 2000 and 2001 (N-new part, P-panel part)

	Year 2000		**Year 2001**			
Cohort	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
1999/Q2	P					
1999/Q3	P	P				
1999/Q4		P	P			
2000/Q1	P		P	P		
2000/Q2	P	P		P	P	
2000/Q3	N	P	P		P	P
2000/Q4		N	P	P		P
2001/Q1			N	P	P	
2001/Q2				N	P	P
2001/Q3					N	P
2001/Q4						N

The above model enabled the 60% overlap between two consecutive quarters and 40% overlap between two consecutive years.

c) Sample selection

The LFS sampling frame was the central register of population combined with the stratification information. The strata were defined with six types of settlements (according to the size and proportion of farmers) and twelve statistical geographical regions. Altogether, due to the process of collapsing small strata, there were 47 strata.

In each stratum the sample selection was a systematic simple random sample. Implicit stratification was additionally performed according to the settlements, streets, and buildings. The sampling rate in each stratum was corrected for the anticipated nonresponse rate. No field substitutions were applied (Vehovar, 1999).

d) Sample size

In each quarter of the year 2000 new units were selected. In addition, around 5,000 (responding) households were included from previous four quarters. Thus, around 7,000 households were included into the sample per quarter and around 6,000 were responding, e.g. 20,000 persons. The quarterly sample size thus represents more than 1% of the population in Slovenia and the yearly database consists of around 80,000 interviews.

1.3. The sampling issues

The LFS sample is restricted to the residential population. The institutional population is thus excluded. The central register of population, which served as a sampling frame, is daily updated and it is a very reliable administrative register of persons with permanent residence in Slovenia. From 2000 it is maintained at the Ministry for Internal Affairs and it successfully serves to all administrative/official purposes (taxes, elections etc.). However, the unit of sample selection is a person and there is no data about his/her relation to the household. Thus, the probability of selection for a household cannot be calculated before data collection; only post-survey adjustments can be applied.

Another problem is the noncoverage arising form eligible households in dwellings where nobody is officially registered. These problems will be in large part solved in the future with the new register of buildings/dwellings and with the new register of households. However, these problems are rather minor, because the noncoverage of the register as the sampling frame is estimated bellow 1%.

1.4. The implementation

a) Interviewing mode

The data are collected with two different modes: CAPI (computer assisted personal interviews) and CATI (computer assisted telephone interviews). All new households from the fresh quarterly sample are interviewed face-to-face. There are around 30 experienced interviewers for LFS all equipped with portable computers. Repeated interviews are made with CATI mode from the telephone center at the Statistical Office, except for the households without telephone and those unable to participate in a telephone interview. However, the fixed telephone coverage rate is already around 95%. The repeated interviews are usually shorter than the initial ones, because most often the responses from previous quarters are simply confirmed. Before interviewing each household gets an advance letter with description of the survey and brochure with LFS results. Beside advance letters there are no other incentives to increase cooperation.

b) Response rates

For a face-to-face interview in the new part of the sample the non-response rates are around 17-18% and refusal rates are around 12-13%. In the repeated

telephone interviewing for the households in the panel, the non-response rates are lower, e.g. around 10-11% and refusal rates around 6-7%.

Standard nonresponse definitions are applied here:

- eligibility rate is the ratio between eligible households and all households included into the sample;
- non-response rate is the ratio between responding households and number of eligible households;
- refusal rate is ratio between number of refusals and number of contacted households;
- non-contacted rate is the ratio between number of non-contacted households and number of eligible households,
- completion rate is the ratio between number of responding households and all households included into the sample.

Table 2. Response rates (%) for LFS 1991-2000 (total sample)

Response rates	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Eligibility rate	97	98	97	97	97	97	95	97	98	95
Response rate	95	91	92	91	90	90	87	88	88	88
Non-response rate	5	9	8	9	10	10	13	12	12	12
Refusal rate	3	5	5	6	8	8	8	9	8	7
Non-contact rate	2	3	1	2	1	2	3	3	3	4
Completion rate	92	89	89	88	88	87	83	85	87	84

From 1991 the non-response rates in LFS have grown considerably, however it seems to stabilize in the last four years at 11-12%. Similarly, the refusal and non-contact rate stabilized at 7-8% and 2-3% respectively.

c) Costs of the survey

In the year 2000 the interviewer's costs were around 86,000 Euros, the travel costs around 74,000 Euros and the mailing costs around 5,000 Euros, which totals to 165,000 Euros. The costs of the statistical staff and other statistical operations are difficult to establish, however, roughly around 3 full time equivalents per year (3 FTE) are involved. In addition, of course, there also exist certain overhead costs.

For the optimization procedures only the variable costs bellow have served as a basis for the calculation of the costs per PSU and the costs per interview.

Table 3. Variable survey costs for LFS 2000 and HBS (in EURO)

Costs of the survey	LFS	HBS

Interviewer costs	86,070	14,640
Travel costs	74,734	28,180
Postal costs	142	3,590
Incentives	-	4,747
Printing costs	4,771	2,990
Total variable cost	**165,717**	**54,148**

d) Weighting

In the first step the data are weighted for unequal probability of selection and for non-response. Next, the post-stratification is performed according to the known population distribution: age (8 groups), sex and region (12 regions). The post-stratification is done on individual level (members of the same household thus receive different weights). The quarterly data are weighted differently compared to the yearly data where all four quarters are represented equally. In a longitudinal analysis the weight from the last quarter is used as a longitudinal weight.

We can observe certain effect of the weights from the following table, representing the categories (expressed in percentages) of the population.

Table 4. Impact of weights on employment status (%), LFS 2000, 2nd Quarter

Working status	Non-weighted	Weighted
Unemployed	3.6	3.5
Employees	38.2	37.7
Self-employed	5.6	5.0
Unpaid family workers	2.8	2.2
Nonactive	49.8	51.6
Total	**100.0**	**100.0**

1.5. Sampling errors and publication criteria

The sampling errors and design effects are routinely estimated only for the key variables: unemployment rate and employment/population ratio. There, the coefficient of variation of the estimates is calculated (CV), too. The ratio estimator – based on administrative data for the population size (e.g. the number of total and active population) – is used for the estimation of the totals. The design effects for large status groups are usually much bellow 2. The unemployment rate has thus the design effect of deff=1.3. Design effect for other variables are not calculated routinely. As an illustration we present here the design effect for the percentage of women. There, the major impact arises from the weights and not from the clustering.

Table 5. Design effect (Deff) for LFS 2000, 2nd Quarter

Variable	Percentage	Deff
Total unemployment rate	7.2%	1.29
Percentage of women	51.2%	1.33

The guidelines from the table bellow are used for the publication of the results.

Table 6. Criteria for the publication of the estimates for LFS 2000

Publication criteria	Coefficient of variation (CV)	The minimal population size of the subgroup - quarterly estimates	The minimal population size of the subgroup yearly estimates
No restriction	< 10%	>10,500	>4,000
( )	10 - 20%	10,500 - 3500	4,000 – 1,000
(( ))	20 - 30%	3,500 - 1,000	1,000 – 500
.	> 30%	< 1000	< 500

Estimates with CV less then 10% are published without any restrictions; estimates with CV between 10-20% are published in a bracket (). With the CV between 20-30% two brackets (()) are applied. When the CV exceeds 30% the results are replaced with a dot "." meaning "nonzero but unreliable".

1.6. Statistical analysis and accessibility of data

Statistical office performs standardized data analysis and tabulation that are included in different national and international statistical publications. The LFS data for the year 2000 were also archived in the national Social Science Data Archive at the Faculty of Social Science. Of course, large and rich data enables much more statistical analysis. In part, researchers from the universities and other institutions perform more detailed studies.

3. Household Budget Survey (HBS)

3.1 The background

The purpose of the HBS survey is to provide representative statistical data on income and expenditure of private households. The results of this survey are particularly important for the following purposes:

- national accounts,
- construction of the weights of the consumer price indices,
- key socio-economic studies (poverty, standard of living etc.).

The first surveys on household consumption took place at the Statistical Office of the Republic of Slovenia in the sixties. Until 1997 the survey was regularly conducted according to the innovative methodology designed by the Federal Statistical Office of Yugoslavia. The sample design was a two-stage sample with stratification of the PSUs at the first stage. Strata were defined according to the region (Ljubljana, settlements outside Ljubljana) and according to the proportion of farmers in enumeration area. Primary sampling units were enumeration areas, sampled with probability proportional to size (PPS). At the second stage, the persons were selected from the central register of population and they also determined the households. In each PSU five households were interviewed. Until 1993 the substitution procedure was used to provide five responding units, however from 1994 the "take" per cluster was increased to 6 to 8 persons within each PSU. Additionally, the corresponding weights were applied.

Until 1997 two different surveys were conducted according to the different reference period: one on a quarterly basis (implemented every year) and another as an annual survey (in five-years intervals). The last annual HBS in 1993 included 3,270 households and the quarterly one included 1,000 households. In the annual survey the interviewing was implemented at the end of the year for the whole year, while with quarterly survey the sampled households were interviewed four times per year.

3.2. The redesign of the HBS sample

The Eurostat puts a considerable effort in harmonizing the HBS surveys in all Member States of the European Union and also in Candidate Countries (Household budget surveys in EU - Methodology and recommendations for harmonization, 1997). Slovenian Statistical Office adopted these guidelines very strictly.

a) Sampling frame

Similar to LFS the register of the population is used to select the adult person. These persons also determine the households. Institutional households are excluded. The discrepancy in the selection probabilities (persons/households) is compensated with weighting.

b) Sample size

Annual sample size includes 1,200 responding households. However, this does not suffice for all purposes. Therefore the "Nordic" model was applied, where data from samples of three consecutive years are joined and recalculated to the middle year. This way a sample size of 3,600 households can be used.

c) Sample design

Similar to LFS the proportional stratification with 47 strata is used also for HBS sample. Due to relatively small sample and high number of strata the stratification is performed only implicitly.

In small settlements (less then 1,0000 inhabitants) the enumeration areas serve as primary sampling units (PSU). They are selected with probability proportional to size (PPS). Four responding households are needed in each PSUs, what was determined with optimization cost-precision calculations based on previous surveys. In larger towns and cities, the simple random sample is applied with systematic selection. As a consequence, the design effects are relatively low, around deff=1.2 for key variables.

The units are selected for each quarter separately and they are allocated into 12 weeks of the corresponding quarter. The thirteenth week is used for the remaining work with nonrespondents. Together with the units (households) the interviewers are also allocated to the corresponding units within each week to optimize the costs and the number of visits in each settlement. On average, in one week an interviewer usually visits up to four new households.

d) Fieldwork

The advance letters are sent one week before the first visit together with the incentive: a pocket calculator. Due to extensive questionnaire and large interviewer burden, there exist a serious potential danger from the response errors. It is thus extremely important to recruit experienced and motivated interviewers for this survey. As this is a continuous survey, it can be implemented with smaller number of the interviewers (e.g. 20). Smaller number of interviewers is also much easier to control.

The interviewers register their contacts/attempts with household in a special form. The status (concerning dwelling, household and reference person) of each unit is thus very clear as well as the number of contacting attempts, number of filled diaries and potential reasons for non-response.

Data are collected with a questionnaire filled by the interviewer and with diaries that are filled by the members of the household. When redesigned surveys started in 1997, a half of interviewers used computers (CAPI) and the other half interviewed with paper questionnaires (PAPI). Today, almost all interviewers conduct interviews with computers. In this way the quality of data increases as well as the efficiency of data processing.

Households keep the diary for 14 days. During this time they regularly fill their daily expenditures.

e) Non-response

No substitutions are used in HBS. Instead, the initial sample size is increased according to the anticipated nonresponse. Of course, the assumption here is that non-respondents do not differ from responding households.

The households are considered as responding if they completed at least the basic interview questionnaire. With more strict criteria the response rates would have been lower. However, two thirds of the data are obtained from the questionnaire, so this is reasonable approach.

In addition to the definitions of the nonresponse rates at the Table 4 the response rate for the completed questionnaires and diaries appear here. This is simply the ratio between households who completed both tasks (questionnaires and diaries) and the number of all eligible households.

Table 7. Sample size and response rates for HBS 1997-2000

Sample size	**1997**	**1998**	**1999**	**2000**
Gross sample size	1666	1684	1624	1620
Completed questionnaires	1276	1301	1290	1268
Completed questionnaires and diaries	1124	1102	1092	1152
Nonresponse rates (%)				
Ineligibility rate	5	5	4	4
Response rate	81	82	82	81
Nonresponse rate	19	18	18	19
Refusal rate	14	12	11	11
Noncontact rate	3	4	5	6
Completion rate	77	77	79	78
Nonresponse structure (%)				
Refusals	69	68	61	56
Noncontact	17	24	29	32
Unable to cooperate	8	4	7	10
Other nonresponse	6	4	3	2
Total	100	100	100	100
Response rate for completed questionnaires and dairies (%)	71	69	70	74

We can observe relatively high response rates (for this type of surveys) at the level of completed interview questioners (around 81%). However, the response rate for complete response from the units (including diary) is considerably lower (around 70%).

f) Imputation

If the first or the second part of the basic interview questionnaire is missing, such a household is treated as nonresponse.

The complete set of all missing diaries for certain household units are compensated with hot deck imputation method based on complete diaries of a similar household (donor).

The item nonresponse is also compensated with hot-deck procedure. Each missing value is replaced with a value of the previous respondent within certain imputation class. Responding households are divided into classes according to household size and other socio-demographic characteristics.

In particular, the missing individual income is substituted with the income of the person with the same employment status and with similar education.

g) Weighting

In HBS weights are applied in the following steps:

- design weights due to unequal probability of selection of persons (households with higher number adults have higher probability of selection) and due to non-response;
- post-stratification weights that provide matching of the distribution of households and persons to the known population values;
- expansion factors needed to compensate different reference periods of the variables. A variety of reference period is finally re-calculated to one year. The coefficient of the recalculation is the ratio between reference period of the survey (e.g. one year) and reference period of the individual variable. Formulas for the calculation of the coefficient are as follows:
  - for diaries: 365 days / number of days of keeping the diary,
  - for questionnaires: 12 months / reference period in months;
- weights needed to deflate the yearly data according to the reference period of the survey. When combining the data for three consecutive years each individual year is taken with the weights recalculated to the reference year in the middle. The calculation for a certain date thus use three-year data with half of the data referring to the period before this date and half to the period after this date.

h) Estimation

In the previous HBS surveys (before 1997) there was a strict balance required between expenditures and incomes. However, due to different methodology of data collection, different reference periods and other changes of the survey design, this property no longer holds. There thus exists a considerable gap between incomes and expenditures. Despite the fact that the questions about income are very detailed the income is often underestimated compared to the expenditures.

In addition, the estimates calculated from three consecutive years to the reference year in the middle are not equal with estimates based on the middle year alone.

Table 8. Comparison of the estimates from annual surveys 1997, 1998, 1999 and from the combined sample 1997-1999, recalculated to the middle year 1998

Estimate	1997	1998	1999	1997-1999
expenditure for clothes (%)	5.7	6.3	5.8	5.9
expenditure for new cars (%)	5.4	5.3	6.7	5.8

We can observe that according to the 1998 yearly estimate 5.3% of total expenditures is spent for new cars, however in the combined three-year data, the 1998 estimate in higher, 5.8%. Therefore, it is very hard to avoid the above types of discrepancies.

For the main groups of variables the standard errors are calculated and expressed as relative precision with coefficients of variation (CV). The design effects are moderate as shown in the table bellow. As expected, in the part with two-stage cluster sample design the values are somehow larger compared to the simple random sample applied in urban areas.

Table 9. Deff for the average available assets of the households, HBS 1997-1999

Variable	Stratified two stage sample	Stratified systematic sample
Available assets	1.50	1.09
Available money assets	1.48	1.08
Income from work under employment	1.42	1.14
Income from work under contract	1.40	1.08
Income from selfemployment	1.31	1.06
Pensions with supplements	1.37	1.15
Other social income	1.51	1.02
Property income	1.16	1.17
Receipts from sale	1.13	0.87
Other receipts	0.92	0.88
Money gifts and transfers	1.48	1.43
Value of own production	1.54	0.98
Food	1.49	1.00
Beverages	1.67	0.87
Heating	1.85	1.00
Benefits in kind	1.47	1.06
From employment	1.48	1.04

g) Costs of the survey

In the year 2000 the interviewer's costs were roughly around 54,000 Euros. The personnel costs are estimated to 3 full-time equivalents per year (3 FTE). The detailed structure of the variable costs was already shown in the Table 3.

h) Presentation of the results

The results of the survey are published in the Statistical Rapid Reports, Statistical Yearbook and some other Slovenian publications. There, similar criteria are applied as with LFS (Table 6). The special series "Results of the survey" bring detailed results of the survey and the methodology. Data also appear in other publications, such as CESTAT Statistical Bulletin as well as in different international publications such as those of the World Bank, UNICEF and Eurostat. Some researchers outside Statistical Office also analyze the micro data.

4. Conclusions

Ten years ago the household sample surveys were rather neglected source of data collection in Slovenian official statistics. However, after a relatively smooth process of the transition, the Slovenian Statistical Office now routinely conducts the standard series of household surveys. At the same time new generations of statisticians were recruited and properly educated.

The basic socio-economic surveys are thus almost completely harmonized with Eurostat requirements (Slovenian Statistical System: A Global Assessment, 2001). This is particularly true for HBS (Household budget surveys in EU - Methodology and recommendations for harmonization, 1997) and LFS (Labour force survey: Methods and definitions, 1998).

Today, besides LFS and HBS, the following household surveys are also being carried out at Statistical Office:

- Household Energy Consumption Survey (HECS - 1997),
- Time Use Survey (TUS), 2000/2001.

In both cases a very similar sample design is applied, e.g. stratified two-stage cluster sample with a "take" 5-10 units per cluster.

Other survey organizations rarely perform full household surveys, because the majority of research target only one persons per household and concentrate on his/her attitudes. On the other hand, due to high telephone penetration, excellent public telephone directory and negligible share of non-listed numbers, all the surveys rapidly move to the telephone. Thus, for example, the household travel survey (Conducted by the Agency for Roads), which included more than 3% of the total population, was performed on the telephone. There, the sampling problem reduces only to the stratification and to the post-survey adjustments.

Among the open issues, the problem of designing samples in a relatively small country should be pointed out. In principle, the request for precision would demand – given the required fixed precision on a national level – basically the same sample sizes for small as well as for large countries. This is, of course, a serious burden for a small country, so the compromising of the design and the sample size is one of the open issues in Slovenian household surveys.

Another aspect for possible improvements is the additional optimization of the cost/precision issues. The rich registered oriented databases (tax registers, database of employees, insurance databases, etc.) that can be linked to

geographical systems and to the Census data offer a certain room for additional improvements of the methodology.

REFERENCES

EUROSTAT (1997), *Household budget surveys in EU - Methodology and recommendations for harmonization 1997*. Office for Official Publications of the European Communities, European Commission, Luxembourg.

HAUSMAN, R., MEHRAN, M., VERMA, V. (1990). *Surveys of economically active population, employment and underemployment*. International Labour Office, Geneva.

EUROSTAT (1998), *Labour force survey Methods and definitions*: Luxembourg.

SLOVENIAN STATISTICAL SYSTEM (2001)*: A Global Assessment*. A Phare Project, Statistical Office of Republic of Slovenia.

VEHOVAR, V., ZALETEL, M. (1995). Nonresponse trends in Slovenia. *Statistics in Transition*, 1995, vol. 2, no. 5, pp. 775-788.

VEHOVAR, V. (1997). The labour force survey in Slovenia. *Statistics in Transition*, June 1997, vol. 3, no. 1, pp. 191-199.

VEHOVAR, V. (1999). Field substitution and unit nonresponse. *Journal of Official Statistics.*, 1999, vol. 15, no. 2, pp. 335-350.

STATISTICS IN TRANSITION, June 2002
Vol. 5, No. 4, pp. 687–700

AN ALTERNATIVE MEASURE OF UNEMPLOYMENT IN TRANSITION ECONOMIES – IDEA OF FREE-MARKET UNEMPLOYMENT

Grzegorz Kuczyński, Krystyna Strzała[1]

ABSTRACT

Economies in transition are subject to privatisation process that may be perceived as a permanent structural shock to the economy and the traditional measures of unemployment seem to be inadequate. That is why we suggest decomposing the observed unemployment rate into two fractions – privatisation and free market unemployment. The former we consider as the level of unemployment directly subject to the transformation process, the latter – as unemployment which formation is subject only to the market forces. In order to reach this aim the Blanchard-Quah decomposition is used, which allows for direct derivation of the privatisation unemployment. The free market unemployment is derived as the difference between the observed and the privatisation unemployment.

Key words: Unemployment, SVAR Modelling, Blanchard-Quah Decomposition JEL Classification: C12, C13, C32, J00

1. Polish labour market and inflation characteristics in period 1990-2000

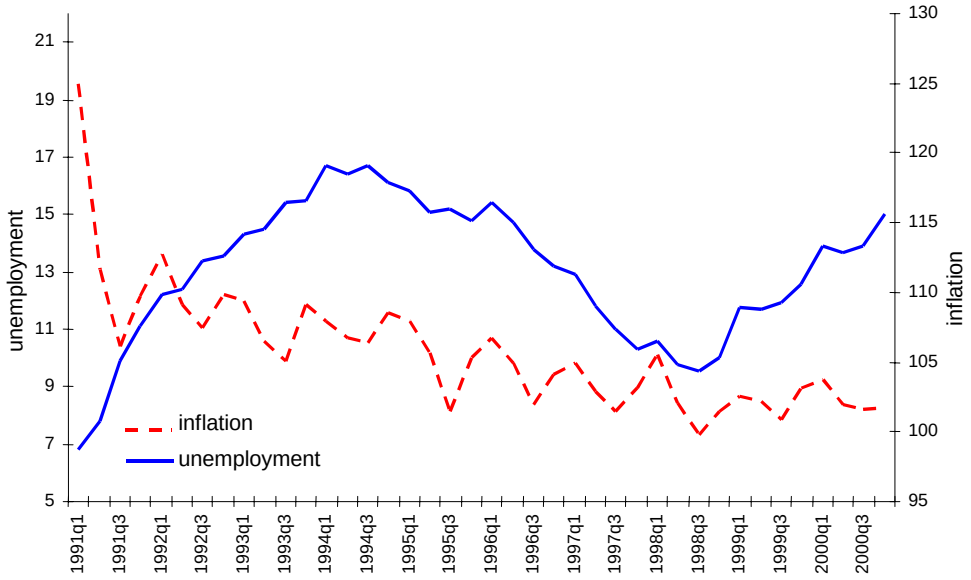
The Phillips Curve theory (see Phillips (1958)) assumes negative relation between unemployment rate and inflation. Figure 1 shows the unemployment rate and inflation in the period 1991q1-2000q4 for Poland. Clear negative relation between unemployment rate and inflation can be observed only in the period 1991q1-1994q1. Later, both variables develop in the same direction. In 1998q3 unemployment begins to rise while inflation still decreases at a very low pace, so the negative relation can again be observed.

Possible explanation of the observed phenomena is the privatisation process, which can be interpreted as a structural shock. Privatisation, especially

[1] Department of Econometrics, University of Gdańsk.

restructuring of the industry, permanently and artificially disturbs the formation of unemployment. Privatisation may influence unemployment in two ways. In the short-run, it may cause an increase in unemployment but a long-run effect of privatisation is a rise in economic activity, hence lower unemployment. The overall result depends on the relation between these trends. Thus, it may prove beneficial to free the official (registered) unemployment rate from the influence of privatisation.

Figure 1. Inflation and unemployment in period 1991q1-2000q4



The proposed approach consists in decomposing the unemployment rate into two fractions:

- free market unemployment rate (un^{fm})
- privatisation unemployment (un^{p})

We assume that free market unemployment is subject to free market processes, i.e. it is the fraction of unemployment that, according to the Phillips Curve theory, is negatively correlated with inflation. Privatisation unemployment, as the fraction of unemployment subject to privatisation process, is independent of inflation in the long-run.

An econometric tool to perform the split of unemployment rate is Blanchard-Quah decomposition (see Blanchard, Quah (1989)). In order to estimate the privatisation unemployment using this approach we assume that in the long-run the cumulated impact of demand shocks on unemployment is equal to zero. The free-market unemployment is estimated as the difference between the registered and the privatisation unemployment.

These remarks are justified by the analysis of the development of the Polish labour market in the last decade:

- The Polish economy entered the transformation period with remarkably high level of hidden unemployment. The phenomenon of unemployment was neither used nor measured in the centrally planned economy. In January 1990, the problem of unemployment emerged. Ex-post estimates of official unemployment rate prove that about 6% of vocationally active population were unemployed (Lavigne (1995)). The registered unemployment rate at the end of December 1991 reached the level of 11,8%, and the number of registered unemployed people equalled 2155,6 thousand. The unemployment rate kept rising dynamically also in subsequent years reaching its highest level of 16,9% at the end of July 1994 when the number of unemployed people was 2982 thousand.

- The transformation process that started at the beginning of 1990 was characterised by a negative demand shock, often referred to as *transformation recession* (see Józefiak (1993), Lavigne (1995)), manifesting itself by a rapid decline of GDP -11,6 % and -7,6 % in years 1990-1991, respectively. The recession also caused a decline in labour force demand and forced changes in the labour market. However it is worth noticing that the reduction of the number of people employed in the national economy in 1990-1991 was significantly smaller than the decline of the real GDP. At the same time, real wages (in USD) in the period 1991-1994 were very stable oscillating around 70% of the 1989 level, after its substantial purchasing power parity decrease (about 24,4%) in 1990.

- It can be concluded (see e.g. Kwiatkowski, Tokarski (1997)) that in the period 1990-1991 hidden unemployment was not decreasing. On the contrary, it was even increasing, because the reduction of the average number of employed persons in the transformation recession period was not so intensive as that of the real GDP. The sold production of the industry in real terms decreased by 24,2% and 11,9% in years 1990 and 1991 respectively. This caused a deep and short-term fall in industry labour productivity. It is probable that in the following years the level of hidden unemployment was gradually decreasing. In the first five years of the transformation process the general fall in average employment in the national economy reached 12,1% of the 1990 level.

- In the period 1995-1998 the growth of real GDP was accompanied by an increase in average employment, average number of employed persons in the national economy and a decrease in unemployment and unemployment rate[1]. However, the average rate of increase in the number of employed persons (2,5%) was less than the average rate of increase in real GDP (6,1%). This

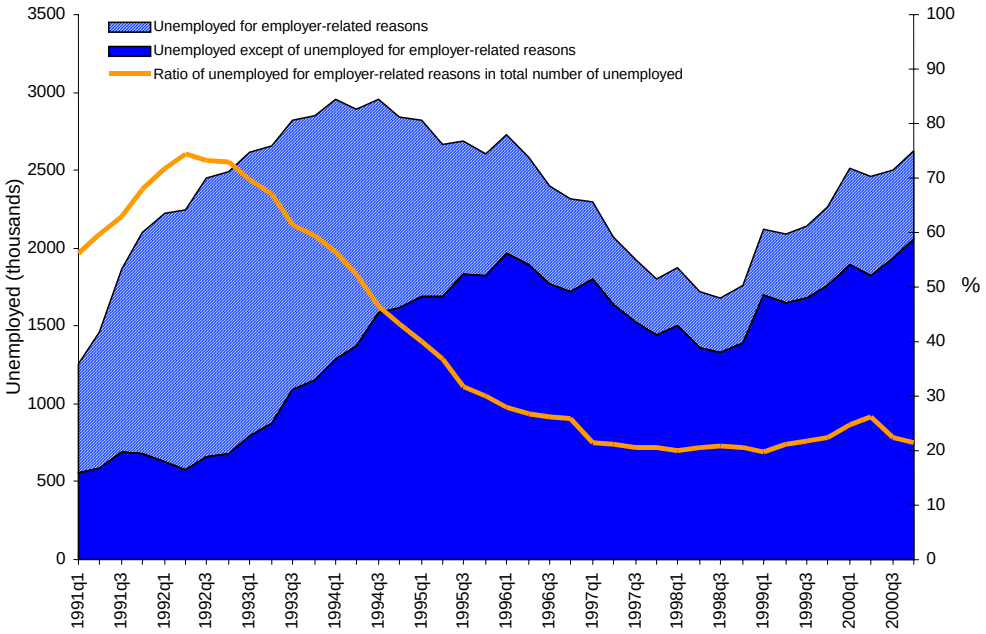
[1] The minimum level of unemployment rate of 9,5% was reached in August 1998.

may contribute to the statement that the period was characterised by a further reduction of hidden unemployment (see Rogut, Tokarski (2000)).

- The transformation process, especially the privatisation of the Polish economy, was conducted in two ways: liquidation and capital manner. As a result of restructuring processes, 700 enterprises were privatised by liquidation manner and 1111 were wound up by the 3rd quarter of 1993. From the labour market perspective, the privatisation process can be regarded as a permanent structural shock, supplying the stock of unemployed with new entrants. In 1991-1994, the ratio of people who lost jobs because of company reasons among new registered unemployed was 63,4% (maximum in 1992 – 73%) and in 1995-1998 this ratio diminished to an average value of 25,8% – from 34,8% in 1995 to 20,6% in 1998. With regard to capital privatisation, no mass employment reduction occurred, because most often the strategic investor was obliged to sign a social package guarantying the specified level of employment in a given period of time.
- The situation in labour market changed after the 4th quarter of 1998. The Russian crisis may be regarded as a negative demand shock (see e.g. Strzała (1999 a, b)) causing drastic reduction of demand for Polish goods, most of all food, with simultaneous increase in imported goods volume that did not reach Russian market. The adjustment process of real economy consists mainly in a decrease in supply and costs optimisation as well as in reduction of employment. As a result, since September 1998 unemployment has been rising, reaching 15,1% at the end of 2000.
- Deregulation of food prices in the middle of 1989 and price liberalisation which covered c.a.90% of prices are regarded as the starting date of transformation in monetary and financial sphere of the economy. Those steps resulted in hyperinflation that emerged in 1990. After the initiation of the stabilisation packet, that among others introduced wage indexation and "progressive tax on over-normative rise of wage fund" in state enterprises, inflation fell, by 250%, from 640% to 70% in 1991. The introduction of nominal "anchor" (indexation) resulted in a rather unplanned by-product – a drastic fall in the purchasing power parity of wages in 1990. Nominal wage indexation indicator set by the government to 0,8 was kept till 1994. To some extent it may explain the relative stability of real wages in that period.
- In 1995, wage controls were changed to a tripartite negotiation system of the government, entrepreneurs and trade unions representatives. In this system, the indicator of maintenance-costs rise played a basic and referential role. This contributed to sustaining cost-push inflation (see Welfe (1996)). It is worth mentioning that adjustment reactions of enterprises, especially national enterprises, to price liberalisation were delayed due to sustained habits typical for period of centrally planned economy. As a result, from the perspective of

the whole economy, the liberalisation process and opening of the economy caused adjustment processes in the real sphere that were distributed in time. These manifested themselves in delayed and distributed in time indispensable reductions of employment.

Figure 2. Unemployed persons in total and for employer-related reason.



2. Methodological remarks

The empirical part of our research focuses on the Blanchard – Quah decomposition, which is applied to extract permanent and transitory components from the unemployment rate series. The first component we interpret as "privatisation" unemployment, while the other one as a "free market" unemployment. The decomposition is based on 2-equation stationary VAR system, where the first equation describes the dynamics of the I(1) variable subject to the decomposition, entering the VAR in first differences. The variable is influenced by two orthogonal shocks, from which one has only a temporary effect on that variable. The second equation explains the I(0) variable[1] that is affected by the same shocks as the I(1) variable (see Blanchard, Quah (1989), Enders (1995)). The decomposition is performed in following steps:

[1] If the second variable is I(1), its first differences can be used.

♦ estimate 2-equation stationary VAR system:

$$\begin{bmatrix} \Delta y_{1t} \\ y_{2t} \end{bmatrix} = A_1 \begin{bmatrix} \Delta y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \ldots + A_p \begin{bmatrix} \Delta y_{1t-p} \\ y_{2t-p} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \tag{1}$$

where: p – maximal lag in the model[1],

y_{1t} – I(1) variable subject to BQ decomposition,
y_{2t} – I(0) variable, influenced by the same shocks as y_{1t},
A_k – matrix of coefficients such that:

$$A_k = \begin{bmatrix} a_{11}(k) & a_{12}(k) \\ a_{21}(k) & a_{22}(k) \end{bmatrix}, \; k=1,...,p,$$

where:

$a_{ij}(k)$ – coefficient on the *j-th* variable lagged by *k* periods, in the *i-th* equation,
e_{1t}, e_{2t} – error terms, with covariance matrix:

$$\Sigma_e = E\left\{ \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} \begin{bmatrix} e_{1t} & e_{2t} \end{bmatrix} \right\} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}, \; t=1,...,T$$

and compute residuals of the above system together with their covariance matrix $\hat{\Sigma}_e$. As it was stressed above, building a VAR system as a basis for the BQ decomposition requires that both variables are subject to the same shocks, so that model [1] can be presented in bivariate moving average form:

$$\begin{aligned} \Delta y_{1t} &= \sum_{k=0}^{\infty} c_{11}(k)\varepsilon_{1t-k} + \sum_{k=0}^{\infty} c_{12}(k)\varepsilon_{2t-k} \\ y_{2t} &= \sum_{k=0}^{\infty} c_{21}(k)\varepsilon_{1t-k} + \sum_{k=0}^{\infty} c_{22}(k)\varepsilon_{2t-k} \end{aligned} \tag{2}$$

where ε_{1t} and ε_{2t} are independent white-noise disturbances with covariance matrix

$$\Sigma_\varepsilon = E\left\{ \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} & \varepsilon_{2t} \end{bmatrix} \right\} = I_2, \; t=1,..T.$$

One of the shocks, e.g. ε_{1t} is regarded as having permanent effect, and ε_{2t} as exerting only temporary influence on y_{1t} variable.

[1] For choosing the number of lags see Lütkepohl (1991), p. 128-132, Enders (1995), p.312-315.

- Compute elements of $C(0)$ matrix which relates residuals from the VAR system (1) and orthogonal shocks ε_{1t} and ε_{2t}:

$$\begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix} = \begin{bmatrix} c_{11}(0) & c_{12}(0) \\ c_{21}(0) & c_{22}(0) \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} = C(0) \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{21t} \end{bmatrix} \tag{3}$$

Relation [3] emerges from equating one -step ahead forecast errors derived for each variable from the VAR system (1) and from its equivalent BMA representation (2).

From (3) it follows, that:

$$\Sigma_e = C(0)\Sigma_\varepsilon C(0)' = C(0)C(0)' \tag{4}$$

Replacing in [4] Σ_e by its estimate $\hat{\Sigma}_e$, calculated in previous step, we get three relations involving four unknown elements of C(0) matrix. The fourth relation, necessary for the unique determination of the $C(0)$ matrix, emerges from assumption that there is no long-run impact of ε_{2t} shock on y_{1t}. (see Enders (1995), p. 334-335 for details).

- Compute

$$\Delta y_{1t}^p = \sum_{k=0}^{\infty} c_{11}(k)\varepsilon_{1t-k} \tag{5}$$

and calculate, by the summation, the permanent component y_{1t}^p. The set of coefficients $c_{11}(k)$ is the orthogonalised impulse response function (describing the instantaneous (k=0) and lagged (k>0) reaction of the y_{1t} to the unit change in ε_{1t}) and can be readily computed using $C(0)$ matrix and estimated A_k matrices (see Lütkepohl (1991), p. 48 and further for details).

3. Empirical Results

Blanchard-Quah decomposition requires that the variable subject to decomposition, i.e. unemployment rate, is I(1). The second (stationary) variable, which undergoes the same orthogonal shocks, is inflation.

To establish order of integration of both series we applied the ADF test (Dickey and Fuller (1979)) together with the KPSS test (Kwiatkowski, Phillips, Schmidt and Shin (1992)) and Leybourne test (Leybourne (1995)). We decided that unambiguous decision could have been taken in conformity of at least two testes: ADF and KPSS or Leybourne and KPSS. The results are given in table 1.

On their basis, we assume that unemployment rate is I(1) and inflation is I(0). Thus, basic prerequisites for BQ decomposition are fulfilled and having 40 observations of unemployment rate and inflation from 1991q1 to 2000q4 we estimated two-equation VAR system by OLS. General form of the model formulated as a basis for the Blanchard - Quah decomposition is as follows:

$$\begin{bmatrix} \Delta unr_t \\ inf_t \end{bmatrix} = A_1 \begin{bmatrix} \Delta unr_{t-1} \\ inf_{t-1} \end{bmatrix} + \ldots + A_p \begin{bmatrix} \Delta unr_{t-p} \\ inf_{t-p} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}, \tag{6}$$

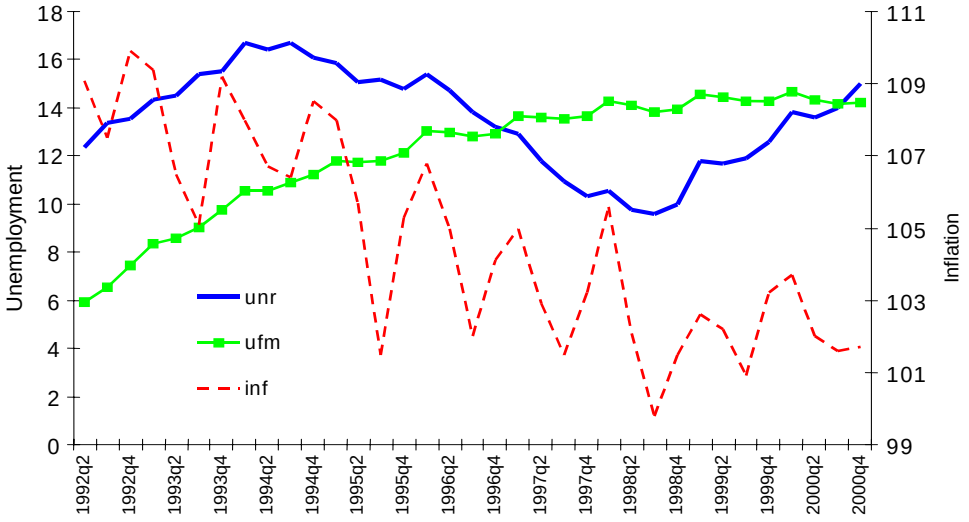
where: unr – unemployment rate (in levels), *inf* – inflation rate (in logs), remaining symbols are the same as in relation (1).

Table 2 presents the results of estimation of lag length choice criteria, assuming that highest admissible order is 4 (forced by the low number of degrees of freedom). General rule for the Schwarz (1978), Hannan-Quinn (1979) and Akaike (1973) criteria is searching for optimal relation between the number of observations and the number of lags. The lag length is chosen according to the minimal value of the given criterion. The results show that all criteria, but the Schwarz criterion (SC), reach the minimal value at lag length equal to 4. Only SC suggests setting lag length to 3. LR test (Lütkepohl (1991), p.125) is also often used for determination of appropriate lag length. Table 3 presents the results of LR test. They confirm the conclusion drawn on the basis of information criteria that optimal lag length is 4. The VAR model, to be correctly specified, should also pass the standard diagnostic tests for white noise processes. In order to verify the general hypothesis of autocorrelation of order *h* in the VAR system, we applied the *portmanteau* test (see Lütkepohl (1991), p.152), as well as Godfrey (1978) serial autocorrelation test for checking residuals in each equation. Table 4 shows the results of Godfrey test for serial autocorrelation. At the 5% significance level there is no autocorrelation in unemployment equation with 4 lags. Inflation equation exhibits no autocorrelation with both 4 and 3 lags. Taking into account the results of the above tests we feel justified selecting 4 lags as the optimal choice.

Thus, having selected the number of lags we estimate the unrestricted version of the model which t-statistics are given in table 5 (top part). As insignificant we treat the parameters with t-statistics smaller than one. In such a case, five parameters are considered as insignificant and restricted to zero. Appropriate restricted model was estimated by the EGLS method (see Lütkepohl (1991), p 169 and further). Table 5 (bottom part) presents the t-statistics of the restricted model. It shows that all but one parameter seem to be significant. The t-statistic referring to the insignificant variable is very close to unity, so we consider the results of estimation of restricted model to be satisfactory. The restricted VAR model is the basis for the BQ decomposition, which we used to recover the privatisation and free market unemployment rate. Performing the

decomposition, we assumed, that the ε_{21t} shock has only a temporary effect on unemployment, so the privatisation unemployment was calculated via summation of the $\Delta un_t^p = \sum_{k=0}^{\infty} c_{11}(k)\varepsilon_{1t-k}$. As a proxy of the initial observation on privatisation unemployment in the first quarter of 1992 we took the fraction of registered unemployment proportional to the ratio of private enterprises to total number of enterprises in that period. Finally, the free-market unemployment was calculated as the difference between the registered and the privatisation unemployment. Figure 3 shows the evolution of the observed (unr) as well as free market unemployment rate (un^{fm}) and inflation (inf).

Figure 3. Inflation, registered, and free-market unemployment rates.



Because of loss of observations in estimation procedure the free market unemployment rate starts from 1992q1. From 1992q1 to 1994q4 the free market unemployment was little lower than the observed one. This proves that in this period the transformation of the Polish economy had negative influence on the level of employment. However from 1997q1, the positive impact of transformation is clearly visible – the observed unemployment rate decreases while the free market one increases. In 1998q2 the difference between both the unemployment rates reaches its maximum value of more than 4,3 percentage points. Since that moment the difference has been decreasing mainly due to a rapid increase in observed unemployment, which on the one hand can be interpreted as the weakening positive effect of transformation on employment,

and on the other hand was due to other non-economic, i.e. mainly regulatory reasons[1].

It is worth mentioning that, whereas the correlation coefficient between registered unemployment rate and inflation equals 0.81, that between inflation and free market unemployment rate is equal to –0.78. These findings can be viewed as a confirmation of the postulated negative relation between inflation and unemployment, although measured in the different way.

In the last quarter of 2000 the registered unemployment rate exceeded that of the free-market unemployment – it is the evidence that unemployment in Poland is really high and under influence of a spectrum of causes, a market forces being only one of them.

4. Concluding remarks

The transformation exerts a very strong impact on the formation of unemployment. We propose a different measure i.e. "free market unemployment rate" which consists in filtering the registered unemployment rate from the influence of privatisation. Our results show that this reasoning proves to give satisfying results. The Blanchard-Quah decomposition allows recovering the unemployment rate consistent with market forces and the Phillips Curve hypothesis.

[1] Health service tax introduced in January 1999 is viewed as a one of the causes increasing registered unemployment.

REFERENCES

AKAIKE, H., (1973) "Information Theory and the Extension of the Maximum Likelihood Principle" w: B.N. Petrov and F. Csaki (eds), *Proceeding of the Second International.Symposium on Information Theory*, Budapest.

BLANCHARD, O.J., D. QUAH, (1989) "The Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review,* 79.

CHAREMZA, W.W., D. F. DEADMAN, (1997) *New directions in Econometric Practice,* Edward Elgar Publishing Ltd., Cheltenham.

CHAREMZA, W.W., E.M. SYCZEWSKA, (1998) "Joint application of the Dickey-Fuller and KPSS tests", *Economic Letters,* 61.

DICKEY, D.A., W.A. FULLER, (1979) "Distributions of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association,* 74.

FRIEDMAN, M., (1968) "The Role of Monetary Policy", *American Economic Review,* 58.

GODFREY, L.G., (1978) "Testing for Higher Order Serial Correlation in Regression Equations when the Regressors Include Lagged Dependent Variables", *Econometrica,* 46.

HAMILTON, J.D., (1994) *Time Series Analysis*, Princeton University Press.

HANNAN, E.J. and B.G. QUINN, (1979) "The Determination of the Order of an Autoregression", *Journal of Royal Statistical Society,* 41.

JOSSA, B., MUSELLA, M., (1998) *Inflation, Unemployment and Money – Interpretations of the Phillips Curve*, Edward Elgar Publishing Ltd.

JÓZEFIAK, C., (1993) „Rachunek mikroekonomiczny a decyzje przedsiębiorstw państwowych w okresie transformacji gospodarki", *Ekonomista,* No 4/1993.

KWIATKOWSKI, D., P.C.B. PHILLIPS, P. SCHMIDT, Y. SHIN, (1992) "Testing the Null Hypothesis of stationarity against the Alternative of a Unit Root"*, Journal of Econometrics,* 73.

KWIATKOWSKI, E., T. TOKARSKI (1997) „Efekty polityki państwa wobec rynku pracy w Polsce.Analiza na podstawie funkcji dostosowań", *Ekonomista***,** No 3/1997.

LAVIGNE, M., (1995) *The Economics of Transition. From Socialist Economy to Market Economy.*, Macmillan Press Ltd., London.

LEYBOURNE, S.J., (1995) "Testing for Unit Roots Using Forward and Reverse Dickey-Fuller Regressions" *Oxford Bulletin of Economics And Statistics,* 57, No 4.

LIPSEY, R.G., (1960) "The Relation between Unemployment and the rate of Change of Money Wage Rates in the United Kingdom 1862-1957: a Further Analysis", *Economica,* 41.

LÜTKEPOHL, H., (1991) *Introduction to Multiple Time Series Analysis,* Springer-Verlag, Berlin.

NELSON, C.R., C.I. PLOSSER, (1982) "Trends Versus Random Walks in Macroeconomic Time Series: Some Evidence and Implications"*, Journal of Monetary Economics,* 10.

PHILLIPS, A. W., (1958) "The Relation between Unemployment and the rate of Change of Money Wage Rates in the United Kingdom 1862-1957", *Economica*, 25.

ROGUT, A., T. TOKARSKI (2000) *Regional Diversity of Employment Structure and Outflows from Unemployment to Employment in Poland*, paper for conference First World EALE/SOLE, Mediolan, 22-25 czerwiec 2000.

SCHWARZ, G., (1978) "Estimating the Dimension of a Model", *Annals of Statistics*, 6.

STRZAŁA K. (red.) (1999a) *Stan i prognoza koniunktury gospodarki Polski*, Instytut Badań nad Gospodarką Rynkową, Biuletyn Nr 23, lipiec 1999, Gdańsk.

STRZAŁA K. (red.) (1999b) *Stan i prognoza koniunktury gospodarki Polski*, Instytut Badań nad Gospodarką Rynkową, Biuletyn Nr 24, październik 1999, Gdańsk.

WELFE, A., (1996) "The Price-Wage Inflationary Spiral in Poland", *Economics of Planning*, 29, No. 1.

Appendix

Table 1. Integration tests.

Deterministic part	ADF		Result	LEYB		Result	KPSS		Result
	I(1)~I(0)	I(2)~I(1)		I(1)~I(0)	I(2)~I(1)		I(0)~I(1)	I(1)~I(2)	
CPI									
No	N/A	N/A	N/A	-1,564	-0,228	N/A			
Constant	-3,458		I(0)	-5,110		I(0)	3,832	0,175	I(1)
Unemployment									
No	N/A	N/A	N/A	-0,227	-1,019	N/A			
Constant	-1,986	-3,103	I(1)	N/A	N/A	N/A	0,578	0,706	N/A

Source: Own computations

Table 2. Information criteria - lag selection

VAR(p)	S.C.	FPE*10^9	AIC	H-Q
4	-9.0659	6513695.50	-9.7769	-9.5315
3	-9.1749	6898738.35	-9.7027	-9.5185
2	-8.2081	21498640.52	-8.5564	-8.4336
1	-7.9149	34177450.53	-8.0873	-8.0260

S.C. -Schwarz criterion, FPE - final prediction error, AIC - Akaike's Information Criterion, HQ - Hannan-Quinn criterion (see Lütkepohl 1991, p. 128-132, Enders 1995, p.312-315). Source: Own computations

Table 3. LR test

Test hypotheses	Test statistic	p-value
$H_0 : p = 4 \quad H_1 : p = 3$	11.5851	0.0207
$H_0 : p = 3 \quad H_1 : p = 2$	51.0819	0.0000
$H_0 : p = 2 \quad H_1 : p = 1$	26.2578	0.0000

Source: Own computations

Table 4. Autocorrelation test.

VAR(p)	χ^2		F	
	Test statistic	p-value	Test statistic	p-value
Unemployment equation				
4	9.2160	0.5851	0.7230	0.0559
3	9.4988	0.5249	0.8188	0.0498
2	12.0000	0.3795	1.0908	0.0174
1	21.8271	0.1846	1.6558	0.0002
Inflation equation				
4	4.0479	0.3045	1.2859	0.3996
3	4.2254	0.2413	1.4656	0.3764
2	24.6371	0.0018	5.6351	0.0001
1	23.7511	0.0000	11.9061	0.0001

Source: Own computations

Table 5. T-statistics – unrestricted model.

VAR(p)	1		2		3		4	
Unrestricted model								
Unemployment	**1.1319**	0.3271	**2.0523**	**1.4957**	0.6415	**1.3216**	**2.2453**	0.6022
Inflation	**1.1393**	**3.4233**	0.5274	**2.3882**	**1.3276**	**3.0447**	0.8151	**1.4124**
Restricted model								
Unemployment	**1.7544**		**2.1635**	**3.5769**		**3.5760**	**2.3190**	
Inflation	**1.0871**	**3.7051**		**2.3266**	0.9331	**3.0833**		**2.7720**

Source: Own computations (the numbers in bold express the significant parameters)

STATISTICS IN TRANSITION, June 2002
Vol. 5, No. 4, pp. 701–720

EMPIRICAL STUDIES OF GENERALIZED CLASSES OF RATIO AND PRODUCT TYPE ESTIMATORS UNDER A LINIEAR MODEL

G.N. Singh[1]

ABSTRACT

In this work, two generalized classes of ratio and product type estimators have been considered for estimating the population mean. Their performances have been examined under a linear model and gamma distributed auxiliary variable. Empirical comparisons are made with sample mean, ratio and product estimators and the recommendations made.

Key words: Ratio and product estimators, linear model, study variable, auxiliary variable, bias, mean squared error.

1. Introduction

Let y_i and x_i be the ith (i = 1, 2,…, n) variates of study variable y and auxiliary variable x respectively in a random sample of size n from a bivariate infinite population with mean $\overline{Y}$ and $\overline{X}$ of respective variates. The sample mean $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ is the natural estimator of $\overline{Y}$. In the presence of an auxiliary variable x, the ratio and product estimators are the most frequently used estimators in sample surveys for estimating the population mean $\overline{Y}$. Their respective use depends upon the nature of correlation (positive or negative) between y and x. The ratio and product estimators are reproduced below as

$$\text{ratio estimator} \qquad \bar{y}_r = \frac{\bar{y}}{\bar{x}}\overline{X} \tag{1}$$

[1] Department of Applied Mathematics, Indian School of Mines, Dhanbad-826 004. INDIA, E. mail: gnsingh_ism@yahoo.com

and

$$\text{product estimator} \qquad \bar{y}_p = \frac{\bar{y}\bar{x}}{\bar{X}} \tag{2}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean of the auxiliary variable x.

In order to improve the efficiencies of ratio and product estimators, ratio-type and product-type estimators may be used. They are developed by mixing the ratio and product estimators with the usual sample mean estimator. Some of the important works in this direction are Ray et-al (1979), Vos (1980), Sahai (1979), Srivastava (1980), Singh et-al (1995), Singh and Singh (1997), Singh et-al (1998) and Singh and Singh (1999) etc. It is to be noted that such estimators generally fall in the most general classes of estimators proposed by Upadhyaya et-al (1985), which are written below

$$T_r = W_{1r}\bar{y} + W_{2r}\bar{y}_r \tag{3}$$

and

$$T_p = W_{1p}\bar{y} + W_{2p}\bar{y}_p \tag{4}$$

where (W_{1r}, W_{2r}) and (W_{1p}, W_{2p}) are suitably chosen scalars whose sums need not be unity.

It has been observed that T_r is more precise than $\bar{y}$ (sample mean estimator) and $\bar{y}_r$ (ratio estimator) and T_p is more efficient than $\bar{y}$ and $\bar{y}_p$ (product estimator) under their respective optimum weights for traditional sampling scheme (design approach) in finite population. Influenced by the wide applicability of T_r and T_p Singh et-al (1993) and Singh (2000) studied the properties of these estimators under super-population approach. In this work, we intend to re-examine the behavior of T_r and T_p under a linear model and gamma distributed auxiliary variates, which was considered by Durbin (1959) and later on by many others.

2. Linear Model

To study the behavior of T_r and T_p, we use a linear model considered by Durbin (1959), which is as follows:

$$y_i = \alpha + \beta x_i + e_i, \quad \beta > 0 \tag{5}$$

$$E(e_i \mid x_i) = 0 \tag{6}$$

$$E(e_i, e_j \mid x_i, x_j) = 0, \quad \forall\ i \neq j \tag{7}$$

$$E(e_i^2 \mid x_i) = V(e_i^2 \mid x_i) = n\,\delta \tag{8}$$

where α and β are unknown real constants and δ is a constant of order n^{-1} and the variates x_i / n have the gamma distribution with parameter h so that $\bar{x}$ has the gamma distribution with parameter m = nh and the density of $\bar{x}$ is written as

$$f(\bar{x}, m) = \frac{1}{\Gamma m} e^{-\bar{x}} \bar{x}^{m-1}, \quad \bar{x} > 0, \ m > 0 \tag{9}$$

3. Biases and Mean Squared Errors (MSE's)

Under the model (5)-(8), T_r and T_p are biased estimators of $\bar{Y}$. The exact expressions of biases and mean squared errors (m. s. e.'s) of T_r and T_p have been derived and produced in the following theorems.

Theorem 1: The exact bias and mean squared error (m. s. e.) of T_r are given by

$$B(T_r) = E(T_r - \bar{Y}) = E\left[W_{1r}(\bar{y} - \bar{Y}) + W_{2r}(\bar{y}_r - \bar{Y}) + W_r \bar{Y}\right] \tag{10}$$

$$= W_{2r}\left(\frac{\alpha}{m-1}\right) + W_r(\alpha + \beta m) \tag{11}$$

where $W_r = W_{1r} + W_{2r} - 1$

and

$$MSE(T_r) = E(T_r - \bar{Y})^2 = E\left[W_{1r}(\bar{y} - \bar{Y}) + W_{2r}(\bar{y}_r - \bar{Y}) + W_r \bar{Y}\right]^2 \tag{12}$$

$$= W_{1r}^2 A_r + W_{2r}^2 B_r + W_r^2 C_r + 2W_{1r}W_{2r}D_r + 2W_{2r}W_r E_r \tag{13}$$

where

$$A_r = V(\bar{y}) = \beta^2 m + \delta \tag{14}$$

$V(\bar{y})$, is the variance of the sample mean estimator $\bar{y}$

$$B_r = MSE(\bar{y}_r) = \frac{\alpha^2(m+2) + m^2\delta}{(m-1)(m-2)} \tag{15}$$

$$C_r = \alpha^2 + \beta^2 m^2 + 2\alpha\beta m \tag{16}$$

$$D_r = \frac{m(\delta - \alpha\beta)}{(m-1)} \tag{17}$$

$$E_r = \frac{\alpha^2 + \alpha\beta m}{(m-1)} \tag{18}$$

Corollary 1: The biases and m.s.e.'s of any other estimators belonging to the class T_r can be easily obtained by putting the suitable choices of their respective weights W_{ir} (i =1,2) in (11) and (13).

Theorem 2: The exact bias and mean squared error (m. s. e.) of T_p are obtained as

$$B(T_p) = E(T_p - \overline{Y}) = E\left[W_{1p}(\bar{y} - \overline{Y}) + W_{2p}(\bar{y}_p - \overline{Y}) + W_p\overline{Y}\right] \tag{19}$$

$$= W_{2p}\beta + W_p(\alpha + \beta m) \tag{20}$$

where $W_p = W_{1p} + W_{2p} - 1$

and

$$MSE(T_p) = E(T_p - \overline{Y})^2 = E\left[W_{1p}(\bar{y} - \overline{Y}) + W_{2p}(\bar{y}_p - \overline{Y}) + W_p\overline{Y}\right]^2 \tag{21}$$

$$= W_{1p}^2 A_p + W_{2p}^2 B_p + W_p^2 C_p + 2W_{1p}W_{2p}D_p + 2W_{2p}W_p E_p \tag{22}$$

$$\text{where} \quad A_p = V(\bar{y}) = \beta^2 m + \delta \tag{23}$$

$$B_p = MSE(\bar{y}_p) =$$

$$\frac{1}{m^2}\left[\alpha^2 m+\beta^2\left\{m(m+1)(m+2)(m+3)-m^3(m+2)\right\}+\delta m(m+1)+4\alpha\beta m(m+1)\right] \quad (24)$$

$$C_p = \alpha^2+\beta^2 m^2+2\alpha\beta m \quad (25)$$

$$D_p = 2\beta^2(m+1)+\alpha\beta+\delta \quad (26)$$

$$E_p = \beta(\alpha+\beta m) \quad (27)$$

Corollary 2: The biases and m.s.e.'s of any other estimators belonging to the class T_p can be easily obtained by putting the suitable choices of their respective weights W_{ip} (i =1,2) in (20) and (22).

4. Optimum Values of W_{ir} and W_{ip} (i=1, 2)

It is seen that T_r, T_p and their respective m.s.e.'s are functions of unknown weights W_{ir} (i=1, 2) and W_{ip} (i=1, 2) respectively. A specific choices of these weights yield a particular member of the classes generated by T_r and T_p. It is, therefore, desirable to detect the members, which have minimum m.s.e.'s. This could be achieved by minimizing MSE (T_r) with respect to W_{ir} (i=1, 2) and MSE (T_p) with respect to W_{ip} (i=1, 2). Since W_{ir} (i=1, 2) are independent to each other, similarly W_{ip} (i=1, 2) are also independent, differentiating MSE (T_r) with respect to W_{ir} (i=1, 2) and MSE (T_p) with respect to W_{ip} (i=1, 2) and equating them to zero, we get the optimum choices of W_{ir} (i=1, 2) and W_{ip} (i=1, 2) as

$$W_{1r}(\text{opt.}) = \frac{C_r(B_r+C_r+2E_r)-(C_r+E_r)(C_r+D_r+E_r)}{(A_r+C_r)(B_r+C_r+2E_r)-(C_r+D_r+E_r)^2} \quad (28)$$

$$W_{2r}(\text{opt.}) = \frac{(A_r+C_r)(C_r+E_r)-C_r(C_r+D_r+E_r)}{(A_r+C_r)(B_r+C_r+2E_r)-(C_r+D_r+E_r)^2} \quad (29)$$

and

$$W_{1p}(\text{opt.}) = \frac{C_p(B_p+C_p+2E_p)-(C_p+E_p)(C_p+D_p+E_p)}{(A_p+C_p)(B_p+C_p+2E_p)-(C_p+D_p+E_p)^2} \quad (30)$$

$$W_{2p}(\text{opt.}) = \frac{(A_p + C_p)(C_p + E_p) - C_p(C_p + D_p + E_p)}{(A_p + C_p)(B_p + C_p + 2E_p) - (C_p + D_p + E_p)^2} \tag{31}$$

5. Efficiency Comparisons-An Empirical Approach

In order to get an insight behavior of the considered (proposed) estimators T_r and T_p under their respective optimality conditions, numerical illustrations are presented. Tables 1-4 present the percent relative efficiencies of T_r over $\bar{y}$ and $\bar{y}_r$ with optimal weights and Tables 5-8 show the percent relative efficiencies of T_p over $\bar{y}$ and $\bar{y}_p$ with optimal weights. The parametric values under the assumed linear model have been considered as: α = 0.0 (0.5) 1.5, β = 0.5 (0.5) 1.5 and δ = 1.0 (1.0) 3.0. The appropriate values of n, h and m are shown in the Tables. In the Tables,

$E_{1r} = [V(\bar{y}) / MSE(T_r)_{opt.}] \times 100.00$, $E_{2r} = [MSE(\bar{y}_r) / MSE(T_r)_{opt.}] \times 100.00$, $E_{1p} = [V(\bar{y}) / MSE(T_p)_{opt.}] \times 100.00$ and $E_{2p} = [MSE(\bar{y}_p) / MSE(T_p)_{opt.}] \times 100.00$.

Since the MSE's of T_r and T_p have been minimized with respect to their weights W_{ir} (i=1, 2) and W_{ip} (i=1, 2), substantial gains over $(\bar{y}, \bar{y}_r)$ and $(\bar{y}, \bar{y}_p)$ are expected, which are apparent from the Tables.

From the values of E_{1r} in Tables 1-4, it is appear that T_r is far better than $\bar{y}$ for smaller values of δ. It is also indicated that the performance of T_r over $\bar{y}$ is improving considerably for fixed values of δ and with the increase in the values of α, β and m.

Table 1. Percent Relative Efficiencies of T_r with respect to $\bar{y}$ and $\bar{y}_r$ for optimum weights when n = 2, h = 4 and m = 8

α	β	δ = 1.0	δ = 2.0	δ = 3.0

		E_{1r} W_{1r}(opt.)	E_{2r} W_{2r}(opt.)	E_{1r} W_{1r}(opt.)	E_{2r} W_{2r}(opt.)	E_{1r} W_{1r}(opt.)	E_{2r} W_{2r}(opt.)
0.0	0.5	224.38	113.97	167.19	127.38	153.49	140.34
		0.1560	0.7604	0.2617	0.5888	0.3353	0.4612
	1.0	611.60	103.55	351.29	107.06	265.97	110.53
		0.0452	0.9318	0.0859	0.8696	0.1227	0.8127
	1.5	1266.62	101.58	676.97	103.16	481.08	104.73
		0.0207	0.9689	0.0404	0.9391	0.0592	0.9105
0.5	0.5	234.77	123.99	170.39	132.36	152.99	141.70
		0.2260	0.6998	0.2988	0.5761	0.3534	0.4776
	1.0	633.60	111.47	363.30	112.88	273.98	115.34
		0.1026	0.8705	0.1346	0.8204	0.1640	0.7740
	1.5	1296.92	108.08	694.73	107.93	493.48	108.82
		0.0632	0.9221	0.0799	0.8965	0.0959	0.8719
1.0	0.5	241.07	141.58	172.90	142.03	153.10	147.27
		0.2931	0.6388	0.3433	0.5485	0.3829	0.4730
	1.0	644.44	126.16	371.66	122.12	280.08	122.46
		0.1552	0.8147	0.1806	0.7739	0.2042	0.7357
	1.5	1305.49	121.06	705.96	115.98	502.55	115.10
		0.1030	0.8787	0.1172	0.8566	0.1308	0.8353
1.5	0.5	244.42	167.80	174.65	156.46	153.34	156.63
		0.3536	0.5831	0.3883	0.5158	0.4168	0.4576
	1.0	646.83	148.20	377.01	135.10	284.55	132.11
		0.2031	0.7642	0.2235	0.7306	0.2425	0.6989
	1.5	1296.97	140.59	711.67	127.51	508.78	123.73
		0.1401	0.8385	0.1522	0.8194	0.1639	0.8008

Table 2. Percent Relative Efficiencies of T_r with respect to $\bar{y}$ and $\bar{y}_r$ for optimum weights when n = 4, h = 4 and m = 16

α	β	$\delta = 1.0$		$\delta = 2.0$		$\delta = 3.0$	
		E_{1r} W_{1r}(opt).	E_{2r} W_{2r}(opt).	E_{1r} W_{1r}(opt).	E_{2r} W_{2r}(opt).	E_{1r} W_{1r}(opt).	E_{2r} W_{2r}(opt).
0.0	0.5	419.89	102.37	257.74	104.73	204.95	107.07
		0.0366	0.9448	0.0704	0.8932	0.1017	0.8450
	1.0	140.83	100.59	747.06	101.19	528.79	101.78
		0.0094	0.9858	0.0187	0.9719	0.0277	0.9582
	1.5	3043.18	100.26	1566.83	100.53	1074.86	100.79
		0.0042	0.9937	0.0084	0.9874	0.0125	0.9812
0.5	0.5	427.34	106.02	261.68	107.27	207.49	109.04
		0.0919	0.8885	0.1189	0.8460	0.1442	0.8059
	1.0	1416.26	103.34	754.60	103.11	534.05	103.40
		0.0407	0.9530	0.0489	0.9404	0.0570	0.9281
	1.5	3061.34	102.64	1577.53	102.10	1082.41	102.10
		0.0257	0.9710	0.0295	0.9651	0.0334	0.9593
1.0	0.5	432.42	112.84	264.72	111.35	209.54	112.04
		0.1419	0.8376	0.1638	0.8023	0.1844	0.7687
	1.0	1423.53	109.26	760.22	106.59	538.30	106.04
		0.0702	0.9220	0.0776	0.9107	0.0848	0.8995
	1.5	3066.92	108.15	1584.69	105.25	1088.21	104.44
		0.0436	0.9493	0.0499	0.9438	0.0534	0.9384
1.5	0.5	435.67	123.02	267.04	117.09	211.18	116.15
		0.1837	0.7916	0.2052	0.7619	0.2220	0.7335
	1.0	1425.56	118.40	764.16	111.69	541.64	109.75
		0.0982	0.8928	0.1047	0.8826	0.1112	0.8724
	1.5	3061.39	116.82	1588.65	109.99	1092.42	107.84
		0.0661	0.9284	0.0694	0.9233	0.0727	0.9182

Table 3. Percent Relative Efficiencies of T_r with respect to $\bar{y}$ and $\bar{y}_r$ for optimum weights when $n = 3$, $h = 8$ and $m = 24$

α	β	$\delta = 1.0$		$\delta = 2.0$		$\delta = 3.0$	
		E_{1r} W_{1r}(opt).	E_{2r} W_{2r}(opt).	E_{1r} W_{1r}(opt).	E_{2r} W_{2r}(opt).	E_{1r} W_{1r}(opt).	E_{2r} W_{2r}(opt).
0.0	0.5	620.60 0.0156	100.92 0.9766	357.86 0.0306	101.84 0.9539	270.81 0.0452	102.76 0.9317
	1.0	2201.24 0.0039	100.23 0.9941	1147.28 0.0078	100.46 0.9882	796.09 0.0117	100.69 0.9824
	1.5	4836.55 0.0018	100.10 0.9974	2464.76 0.0035	100.02 0.9947	1674.23 0.0052	100.31 0.9921
0.5	0.5	625.91 0.0551	102.93 0.9361	360.76 0.0681	103.25 0.9161	272.81 0.0807	103.91 0.8967
	1.0	2210.71 0.0249	101.80 0.9723	1152.58 0.0286	101.49 0.9668	799.79 0.0321	101.54 0.9614
	1.5	4849.31 0.0160	101.50 0.9826	2472.27 0.0176	101.08 0.9801	1679.52 0.0193	101.00 0.9776
1.0	0.5	629.74 0.0918	107.03 0.8986	363.11 0.1030	105.67 0.8810	274.49 0.1139	105.72 0.8638
	1.0	2216.00 0.0451	105.46 0.9514	1156.64 0.0485	103.57 0.9463	802.86 0.0518	103.08 0.9412
	1.5	4853.34 0.0298	104.98 0.9682	2477.40 0.0314	102.99 0.9659	1683.66 0.0329	102.39 0.9635
1.5	0.5	632.34 0.1260	113.27 0.8638	364.98 0.1357	109.14 0.8482	275.89 0.1452	108.23 0.8329
	1.0	2217.57 0.0645	111.23 0.9313	1159.59 0.0676	106.70 0.9265	805.35 0.0707	105.31 0.9218
	1.5	4849.34 0.0433	110.56 0.9543	2480.31 0.0448	105.96 0.9520	1686.73 0.0462	104.48 0.9497

Table 4. Percent Relative Efficiencies of T_r with respect to $\bar{y}$ and $\bar{y}_r$ for optimum weights when n = 8, h = 4 and m = 32

α	β	δ = 1.0		δ = 2.0		δ = 3.0	
		E_{1r} W_{1r}(opt).	E_{2r} W_{2r}(opt).	E_{1r} W_{1r}(opt).	E_{2r} W_{2r}(opt).	E_{1r} W_{1r}(opt).	E_{2r} W_{2r}(opt).
0.0	0.5	821.34	100.48	458.49	100.97	337.84	101.45
		0.0085	0.9872	0.0169	0.9746	0.0251	0.9621
	1.0	3000.70	100.12	1547.68	100.24	1063.41	100.36
		0.0021	0.9968	0.0043	0.9936	0.0064	0.9904
	1.5	6633.45	100.05	3363.96	100.11	2274.17	100.16
		0.0010	0.9986	0.0019	0.9971	0.0029	0.9957
0.5	0.5	825.43	101.82	460.73	101.88	339.40	102.20
		0.0389	0.9562	0.0464	0.9447	0.0538	0.9334
	1.0	3007.98	101.20	1551.75	100.92	1066.25	100.91
		0.0179	0.9806	0.0199	0.9775	0.0219	0.9745
	1.5	6643.26	101.03	3369.73	100.70	2278.23	100.62
		0.0116	0.9876	0.0125	0.9863	0.0134	0.9849
1.0	0.5	828.46	104.72	462.59	103.56	340.74	103.45
		0.0675	0.9269	0.0743	0.9165	0.0809	0.9062
	1.0	3012.12	103.84	1554.91	102.38	1068.63	101.97
		0.0332	0.9649	0.0351	0.9619	0.0370	0.9590
	1.5	6646.41	103.58	3373.71	102.06	2281.44	101.59
		0.0220	0.9769	0.0228	0.9756	0.0237	0.9743
1.5	0.5	830.59	109.21	464.12	106.02	341.89	105.22
		0.0946	0.8994	0.1006	0.8899	0.1066	0.8805
	1.0	3013.39	108.06	1557.25	104.63	1070.60	103.56
		0.0480	0.9496	0.0498	0.9468	0.0516	0.9441
	1.5	6643.27	107.69	3376.01	104.22	2283.85	103.09
		0.0321	0.9665	0.0330	0.9652	0.0338	0.9639

Table 5. Percent Relative Efficiencies of T_p with respect to $\bar{y}$ and $\bar{y}_p$ for optimum weights when n = 2, h = 4 and m = 8

α	β	δ = 1.0		δ = 2.0		δ = 3.0	
		E_{1p} W_{1p}(opt).	E_{2p} W_{2p}(opt).	E_{1p} W_{1p}(opt).	E_{2p} W_{2p}(opt).	E_{1p} W_{1p}(opt).	E_{2p} W_{2p}(opt).
0.0	0.5	205.80	827.45	166.11	550.93	156.25	447.27
		1.5807	-0.5971	1.4173	-0.5039	1.2800	-0.4267
	1.0	346.58	1728.09	268.75	1236.25	229.26	982.19
		1.7248	-0.6803	1.6744	-0.6512	1.6264	-0.6235
	1.5	432.89	2268.41	360.04	1812.59	313.52	1520.02
		1.7539	-0.6972	1.7306	-0.6837	1.7077	-0.6704
0.5	0.5	214.76	946.27	168.91	605.72	154.73	478.70
		1.5729	-0.5777	1.4441	-0.5049	1.3319	-0.4423
	1.0	366.00	1917.68	279.22	1348.12	235.92	1059.62
		1.7040	-0.6604	1.6604	-0.6353	1.6187	-0.6113
	1.5	454.12	2461.07	374.19	1947.54	323.80	1622.39
		1.7374	-0.6823	1.7163	-0.6701	1.6956	-0.6581
1.0	0.5	233.09	1073.60	171.18	666.01	154.36	515.19
		1.5548	-0.5533	1.4521	-0.4959	1.3602	-0.4452
	1.0	385.08	2117.96	289.34	1464.79	242.34	1140.11
		1.6826	-0.6403	1.6447	-0.6186	1.6081	-0.5977
	1.5	475.49	2663.70	388.23	2087.93	333.93	1728.26
		1.7208	-0.6674	1.7016	-0.6574	1.6829	-0.6456
1.5	0.5	230.62	1208.36	173.54	730.76	154.59	555.56
		1.5323	-0.5274	1.4493	-0.4817	1.3735	-0.4404
	1.0	403.74	2328.54	299.06	1585.97	248.50	1223.42
		1.6612	-0.6206	1.6279	-0.6017	1.5958	-0.5834
	1.5	496.96	2876.31	402.12	2233.64	343.86	1837.51
		1.7043	-0.6528	1.6869	-0.6428	1.6698	-0.6330

Table 6. Percent Relative Efficiencies of T_p with respect to $\bar{y}$ and $\bar{y}_p$ for optimum weights when n = 4, h = 4 and m = 16

α	β	δ = 1.0		δ = 2.0		δ = 3.0	
		E_{1p} W_{1p}(opt).	E_{2p} W_{2p}(opt).	E_{1p} W_{1p}(opt).	E_{2p} W_{2p}(opt).	E_{1p} W_{1p}(opt).	E_{2p} W_{2p}(opt).
0.0	0.5	326.15	1298.48	230.38	805.11	192.56	606.06
		1.8253	-0.7993	1.7710	-0.7640	1.7193	-0.7304
	1.0	606.10	2725.24	458.59	1974.50	377.43	1560.61
		1.8678	-0.8271	1.8535	-0.8177	1.8393	-0.8085
	1.5	755.19	3483.14	630.27	2848.13	545.17	2415.24
		1.8759	-0.8324	1.8694	-0.8281	1.8630	-0.8240
0.5	0.5	337.54	1416.61	235.35	864.77	195.42	645.14
		1.7913	-0.7667	1.7446	-0.7366	1.6999	-0.7079
	1.0	627.24	2899.26	470.53	2081.85	385.37	1636.88
		1.8473	-0.8083	1.8341	-0.7997	1.8210	-0.7912
	1.5	777.21	3652.02	645.55	2971.60	556.55	2511.38
		1.8617	-0.8194	1.8556	-0.8155	1.8495	-0.8115
1.0	0.5	348.06	1538.01	239.90	925.85	198.05	685.22
		1.7590	-0.7358	1.7183	-0.7101	1.6798	-0.6853
	1.0	648.14	3078.67	482.15	2191.45	393.05	1714.42
		1.8276	-0.7902	1.8153	-0.7822	1.8032	-0.7744
	1.5	799.31	3825.76	660.75	3097.79	567.80	2609.25
		1.8479	-0.8069	1.8421	-0.8031	1.8363	-0.7994
1.5	0.5	357.77	1662.53	244.05	988.29	200.47	726.25
		1.7286	-0.7069	1.6934	-0.6846	1.6594	-0.6632
	1.0	668.78	3263.39	493.46	2303.24	400.46	1793.19
		1.8086	-0.7727	1.7972	-0.7654	1.7860	-0.7582
	1.5	821.48	4004.38	675.85	3226.69	578.92	2708.81
		1.8344	-0.7946	1.8289	-0.7910	1.8235	-0.7875

Table 7. Percent Relative Efficiencies of T_p with respect to $\bar{y}$ and $\bar{y}_p$ for optimum weights when n = 3, h = 8 and m = 24

α	β	δ = 1.0		δ = 2.0		δ = 3.0	
		E_{1p} W_{1p}(opt).	E_{2p} W_{2p}(opt).	E_{1p} W_{1p}(opt).	E_{2p} W_{2p}(opt).	E_{1p} W_{1p}(opt).	E_{2p} W_{2p}(opt).
0.0	0.5	455.92	1814.19	306.58	1107.37	245.52	816.69
		1.8937	-0.8682	1.8680	-0.8506	1.8428	-0.8335
	1.0	870.62	3771.22	655.56	2756.69	534.01	2183.03
		1.9134	-0.8816	1.9068	-0.8771	1.9002	-0.8726
	1.5	1080.93	4763.05	905.17	3934.21	782.64	3356.26
		1.9171	-0.8841	1.9141	-0.8821	1.9112	-0.8801
0.5	0.5	467.50	1930.53	311.63	1166.58	248.56	855.87
		1.8641	-0.8401	1.8410	-0.8245	1.8184	-0.8092
	1.0	892.18	3939.35	667.77	2861.82	542.10	2258.13
		1.8973	-0.8666	1.8911	-0.8624	1.8849	-0.8582
	1.5	1103.11	4923.67	920.73	4053.38	794.28	3449.85
		1.9062	-0.8740	1.9033	-0.8721	1.9005	-0.8701
1.0	0.5	478.42	2048.95	316.33	1226.59	251.38	895.55
		1.8362	-0.8136	1.8154	-0.7996	1.7949	-0.7859
	1.0	913.58	4111.12	679.75	2968.47	549.99	2334.06
		1.8818	-0.8521	1.8759	-0.8481	1.8700	-0.8441
	1.5	1125.35	5087.50	936.23	4174.38	805.82	3544.60
		1.8955	-0.8640	1.8928	-0.8622	1.8900	-0.8603
1.5	0.5	488.71	2169.38	320.70	1287.39	254.00	935.75
		1.8099	-0.7886	1.7910	-0.7760	1.7725	-0.7636
	1.0	934.81	4286.49	691.50	3076.57	557.68	2410.79
		1.8667	-0.8380	1.8611	-0.8342	1.8555	-0.8304
	1.5	1147.64	5254.57	951.67	4297.20	817.26	3640.50
		1.8851	-0.8543	1.8824	-0.8525	1.8798	-0.8507

Table 8. Percent Relative Efficiencies of T_p with respect to $\bar{y}$ and $\bar{y}_p$ for optimum weights when n = 8, h = 4 and m = 32

α	β	δ = 1.0		δ = 2.0		δ = 3.0	
		E_{1p} W_{1p}(opt).	E_{2p} W_{2p}(opt).	E_{1p} W_{1p}(opt).	E_{2p} W_{2p}(opt).	E_{1p} W_{1p}(opt).	E_{2p} W_{2p}(opt).
0.0	0.5	587.62	2339.26	385.05	1419.27	301.13	1037.28
		1.9244	-0.9022	1.9095	-0.8918	1.8948	-0.8816
	1.0	1136.25	4827.99	854.10	3548.30	692.41	2814.76
		1.9356	-0.9100	1.9319	-0.9074	1.9281	-0.9048
	1.5	1407.47	6057.95	1181.13	5031.53	1021.43	4307.24
		1.9377	-0.9115	1.9360	-0.9103	1.9344	-0.9091
0.5	0.5	599.19	2454.48	389.97	1477.94	304.07	1076.14
		1.8997	-0.8786	1.8861	-0.8691	1.8725	-0.8598
	1.0	1158.00	4993.06	866.42	3652.22	700.52	2889.18
		1.9227	-0.8978	1.9191	-0.8952	1.9155	-0.8927
	1.5	1429.70	6214.40	1196.81	5148.47	1033.17	4399.49
		1.9290	-0.9032	1.9274	-0.9021	1.9257	-0.9010
1.0	0.5	610.23	2571.24	394.62	1537.17	306.84	1115.35
		1.8763	-0.8562	1.8637	-0.8475	1.8512	-0.8389
	1.0	1179.63	5160.89	878.56	3757.28	708.47	2964.19
		1.9101	-0.8858	1.9066	-0.8834	1.9032	-0.8810
	1.5	1451.99	6373.26	1212.44	5266.80	1044.85	4492.62
		1.9204	-0.8951	1.9188	-0.8940	1.9173	-0.8929
1.5	0.5	620.78	2689.49	399.09	1596.96	309.44	1154.91
		1.8540	-0.8348	1.8423	-0.8268	1.8307	-0.8189
	1.0	1201.13	5331.44	890.51	3863.45	716.27	3039.82
		1.8978	-0.8742	1.8945	-0.8719	1.8912	-0.8696
	1.5	1474.31	6534.52	1228.04	5386.53	1056.44	4586.60
		1.9120	-0.8872	1.9105	-0.8861	1.9089	-0.8850

Table 9. Percent Relative Efficiencies of T_r and T_p with respect to $(\bar{y}, \bar{y}_r)$ and $(\bar{y}, \bar{y}_p)$ respectively for different weights when n = 2, h = 4, $\alpha = 0.5$, $\beta = 0.5$, and $\delta = 1.0$

W_{1r} & W_{1p}	W_{2r} & W_{2p}							
	0.3		0.5		0.7		0.9	
	E'_{1r}	E'_{2r}	E'_{1r}	E'_{2r}	E'_{1r}	E'_{2r}	E'_{1r}	E'_{2r}
	E'_{1p}	E'_{2p}	E'_{1p}	E'_{2p}	E'_{1p}	E'_{2p}	E'_{1p}	E'_{2p}
0.01	-	-	-	-	127.08	-	212.74	112.28
	-	133.07	-	185.60	-	179.91	-	124.58
0.04	-	-	-	-	144.19	-	216.26	114.14
	-	142.71	-	194.70	-	179.75	-	120.65
0.07	-	-	-	-	162.84	-	213.36	112.61
	-	153.13	-	203.41	-	178.58	-	116.54
0.10	-	-	-	-	182.28	-	204.53	107.95
	-	164.32	-	211.51	-	176.43	-	112.29
0.13	-	-	-	-	201.17	106.17	191.16	100.89
	-	176.28	-	218.76	-	173.38	-	107.97
0.16	-	-	105.58	-	217.62	114.86	175.05	-
	-	188.99	-	224.91	-	169.52	-	103.63
0.19	-	-	119.36	-	229.42	121.08	157.89	-
	-	202.37	-	229.73	-	164.96	-	-
0.22	-	-	134.70	-	234.63	123.83	140.97	-
	-	216.29	-	233.03	-	159.82	-	-
0.25	-	-	151.29	-	232.30	122.60	125.10	-
	-	230.58	-	234.67	-	154.23	-	-
0.28	-	-	168.50	-	222.86	117.62	110.70	-
	-	244.97	-	234.58	-	148.33	-	-

Note: Dashes in the table indicate that there is no gain with the corresponding choice of weights

Table 10. Percent Relative Efficiencies of T_r and T_p with respect to $(\bar{y}, \bar{y}_r)$ and $(\bar{y}, \bar{y}_p)$ respectively for different weights when n = 3, h = 8, α = 1.0, β = 1.5, and δ = 2.0

W_{1r} & W_{1p}	W_{2r} & W_{2p}							
	0.3		0.5		0.7		0.9	
	E'_{1r}	E'_{2r}	E'_{1r}	E'_{2r}	E'_{1r}	E'_{2r}	E'_{1r}	E'_{2r}
	E'_{1p}	E'_{2p}	E'_{1p}	E'_{2p}	E'_{1p}	E'_{2p}	E'_{1p}	E'_{2p}
0.01	-	-	-	-	-	-	439.94	-
	-	-	-	-	-	115.29	-	120.95
0.04	-	-	-	-	-	-	827.48	-
	-	-	-	-	-	124.05	-	119.11
0.07	-	-	-	-	-	-	1661.00	-
	-	-	-	-	-	132.43	-	115.93
0.10	-	-	-	-	-	-	2205.00	-
	-	-	-	-	-	139.99	-	111.62
0.13	-	-	-	-	135.39	-	1310.50	-
	-	-	-	101.78	-	146.22	-	106.45
0.16	-	-	-	-	191.73	-	653.31	-
	-	-	-	112.57	-	150.67	-	100.68
0.19	-	-	-	-	285.45	-	362.81	-
	-	-	-	124.32	-	152.96	-	-
0.22	-	-	-	-	444.68	-	225.23	-
	-	-	-	136.86	-	152.87	-	-
0.25	-	-	-	-	688.53	-	151.94	-
	-	-	-	149.86	-	150.43	-	-
0.28	-	-	-	-	897.69	-	108.92	-
	-	-	-	162.83	-	145.84	-	-

Note: Dashes in the table indicate that there is no gain with the corresponding choice of weights.

As far as the efficiency of T_r under optimality condition with respect to $\bar{y}_r$ is concerned, we observe that T_r is performing well for the smaller values of m and with the increase in δ, while for fixed values of δ and m efficiency of T_r over $\bar{y}_r$ is increasing with the increase in α and for fixed α decreasing trend could be noticed over the increase in β.

From Tables 5-8, it is observed that T_p is performing well over $\bar{y}$ and $\bar{y}_p$ with the increase in the values of α, β, m and for the fixed values of δ, while, T_p is loosing the gain in performance over $\bar{y}$ and $\bar{y}_p$ with the increase in δ and for the fixed values of α, β and m.

The percent relative efficiencies of T_r and T_p defined over $(\bar{y}, \bar{y}_r)$ and $(\bar{y}, \bar{y}_p)$ respectively are also computed with respect to different values of W_{ir} (i=1, 2) and W_{ip} (i=1, 2) and are expressed as

$E'_{1r} = [V(\bar{y}) / MSE(T_r)] \times 100.00$, $E'_{2r} = [MSE(\bar{y}_r) / MSE(T_r)] \times 100.00$,

$E'_{1p} = [V(\bar{y}) / MSE(T_p)] \times 100.00$ and $E'_{2p} = [MSE(\bar{y}_p) / MSE(T_p)] \times 100.00$.

The values of E'_{1r}, E'_{2r}, E'_{1p} and E'_{2p} are compiled for some selected values of parameters and shown in Tables 9-12. From these Tables, it may be observed that the use of proposed estimators T_r and T_p are limited for some specific weights. However, the uses of proposed classes of estimators T_r and T_p are recommended in preference over $\bar{y}, \bar{y}_r$ and $\bar{y}_p$ or any other estimators belonging to these classes when the optimum choices of weights are considered.

5. Concluding Remarks

In the present work, the properties of more general classes of ratio and product type estimators, which include a number of existing estimators as special cases are discussed. It was observed that the proposed classes work efficiently under a linear model and are preferable over conventional sample mean, ratio and product estimators under optimum choices of weights but not performing uniformly well for specific choices of weights. The performances of T_r and T_p have been studied in design and super-population approach by some authors and the proposition that above are the wider classes with more efficiencies have been justified in this work.

REFERENCES

DURBIN, J. (1959): A note on the application of Quenouille's method of bias reduction in estimation of ratios. Biometrika, 46, 477-480.

RAY, S. K., SAHAI, A. and SAHAI, A. (1979): A note on ratio and product type estimators. Annals of the Institute of Mathematical Statistics, 31, 141-144.

SAHAI, A. (1979): An efficient variant of the product and ratio estimators. Statistica Neerlandica, 33, 27-35.

SINGH, V. K., SINGH, G. N. and SHUKLA, D. (1993): A general class of product-type estimators under super-population model. Jour. Ind. Soc. Ag. Statistics, 45, 2, 177-186.

SINGH, S., MANGAT, N. S. and MAHAJAN, P. K. (1995): General classes of estimators. Jour. Ind. Soc. Ag. Statistics, 47, 2, 129-133.

SINGH, A. K. and SINGH, H. P. (1997): A note on the efficiencies of three product type estimators under a linear model. Jour. Ind. Soc. Ag. Statistics, 50, 2, 130-134.

SINGH, R., SINGH, H. P. and ESPEJO, M. R. (1998): The efficiency of an alternative to ratio estimator under a super-population model. Jour. of Statistical Planning and Inference, 71, 287-301.

SINGH, A. K. and SINGH, H. P. (1999): Efficiency of a class of unbiased ratio-type estimators under a linear model. Jour. Ind. Soc. Ag. Statistics, 52, 2, 228-236.

SINGH, G. N. (2000): A general class of ratio-type estimators under super-population model. Biometrical Journal, 42, 3, 363-375.

SRIVASTAVA, S. K. (1980): A class of estimators using auxiliary information in sample surveys. Canadian Journal of Statistics, 8, 253-254.

UPADHYAYA, L.N., SINGH, H. P. and VOS, J. W. E. (1985): On the estimation of population means and ratios using supplementary information. Statistica Neerlandica, 39, 309-318.

VOS, J. W. E. (1980): Mixing of direct, ratio and product method estimators. Statistica Neerlandica, 34, 209-218.

Table 11. Percent Relative Efficiencies of T_r and T_p with respect to $(\bar{y}, \bar{y}_r)$ and $(\bar{y}, \bar{y}_p)$ respectively for different weights when n=2, h= 4, $\alpha = 0.5$, $\beta = 0.5$, and $\delta = 1.0$

W_{1r} & W_{1p}	W_{2r} & W_{2p}											
	-0.40		-0.50		-0.60		-0.70		-0.80		-0.90	
	E'_{1r}	E'_{1p}	E'_{1r}	E'_{1p}	E'_{1r}	E'_{1p}	E'_{1r}	E'_{1p}	E'_{1r}	E'_{1p}	E'_{1r}	E'_{1p}
	E'_{2r}	E'_{2p}	E'_{2r}	E'_{2p}	E'_{2r}	E'_{2p}	E'_{2r}	E'_{2p}	E'_{2r}	E'_{2p}	E'_{2r}	E'_{2p}
1.00	-	-	-	-	-	-	-	-	-	-	-	-
	-	295.39	-	196.06	-	135.32	-	-	-	-	-	-
1.20	-	143.88	-	-	-	-	-	-	-	-	-	-
	-	633.99	-	410.24	-	258.24	-	170.02	-	118.13	-	-
1.40	-	192.93	-	192.04	-	128.82	-	-	-	-	-	-
	-	850.08	-	846.17	-	567.63	-	342.93	-	215.29	-	143.92
1.60	-	103.99	-	170.24	-	214.48	-	167.16	-	101.69	-	-
	-	458.19	-	750.13	-	945.04	-	736.55	-	448.09	-	271.73
1.80	-	-	-	-	-	128.82	-	188.12	-	185.19	-	124.76
	-	217.59	-	345.84	-	567.63	-	828.92	-	815.97	-	549.71
2.00	-	-	-	-	-	-	-	-	-	139.53	-	166.23
	-	118.82	-	170.74	-	258.24	-	406.15	-	614.83	-	732.47

Note: Dashes in the table indicate that there is no gain with the corresponding choice of weights.

Table 12. Percent Relative Efficiencies of T_r and T_p with respect to $(\bar{y}, \bar{y}_r)$ and $(\bar{y}, \bar{y}_p)$ respectively for different weights when n =3, h = 8, α = 1.0, β = 1.5, and δ = 2.0

W_{1r} & W_{1p}	W_{2r} & W_{2p}											
	-0.40		-0.50		-0.60		-0.70		-0.80		-0.90	
	E'_{1r}	E'_{1p}	E'_{1r}	E'_{1p}	E'_{1r}	E'_{1p}	E'_{1r}	E'_{1p}	E'_{1r}	E'_{1p}	E'_{1r}	E'_{1p}
	E'_{2r}	E'_{2p}	E'_{2r}	E'_{2p}	E'_{2r}	E'_{2p}	E'_{2r}	E'_{2p}	E'_{2r}	E'_{2p}	E'_{2r}	E'_{2p}
1.00	-	-	-	-	-	-	-	-	-	-	-	-
	-	140.06	-	-	-		-		-		-	
1.20	-	-	-	-	-		-		-		-	
	-	342.65	-	173.73	-	100.21	-		-		-	
1.40	-	282.97	-	192.96	-	-	-		-		-	
	-	1261.69	-	860.34	-	342.75	-	165.06	-		-	
1.60	-	-	-	197.73	-	531.94	-	204.93	-		-	
	-	309.36	-	881.61	-	2371.76	-	913.71	-	317.18	-	151.60
1.80	-	-	-	-	-	-	-	322.08	-	782.12	-	177.63
	-	-	-	176.31	-	401.38	-	1463.07	-	3487.25	-	791.98
2.00	-	-	-	-	-	-	-	-	-	112.90	-	453.56
	-	-	-	-	-	109.57	-	205.59	-	503.40	-	2022.27

Note: Dashes in the table indicate that there is no gain with the corresponding choice of weights.