Romania 1997: Survey Information

Summary table

Generic information	
Name of survey	Romania Integrated Household Survey (RIHS) / *Ancheta Integrata în Gospodarii*
Institution responsible	National Institute of Statistics
Frequency	Annual (since 1995)
Survey year / Wave	1997
Collection period	January 1997 to December 1997
Survey structure	Cross-sectional (1/6 of the sample is longitudinal)
Coverage	Nationally representative
Geographic information	Area of residence, region, county, locality (only first 2 delivered to LIS)
Files delivered	Pre-lissified files (one at the household level and one at the and personal level) (there are over 200 original files broken down by sections of the questionnaire)
Sample size	
Households	32,187 interviewed households
Individuals	92,334 individuals
Sampling	
Sampling design	Two-stage sampling with small areas (Census sectors) selected randomly in a first stage and 20 households selected within each area in the second stage.
Sampling frame	Census of January 1992
Questionnaires	Household Questionnaire and Household Income and Expenses Diary
Standard classifications	
Education	Own classification (8 categories)
Occupation	1-digit ISCO-88 code
Industry	Own classification (15 categories)
Income	
Reference period	Each household has a different calendar month as reference period according to the interview date (from January 1997 to December 1997)
Unit of collection	Household (wage income, pensions and unemployment benefits are also collected at the individual level, while self-employment income is also collected at the activity level).
Period of collection	Monthly income.
Gross/net	Variables are collected gross of taxes and contributions, which are also collected separately.
Data editing / processing	
Consistency checks	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.
Weighting	The weighting process comprised of the following three steps: calculation of the basic weights, non-response adjustment, and final adjustment of the samp le population and calculation of the extension coefficients.
Imputation	There are automatic imputing procedures based on logical correlation's, so that there are no missing data.

This document draws extensively from the following documents: "Romanian Integrated Household Survey 1994 – Basic Information" World Bank, August 1998, "Romanian Integrated Household Survey 1997 – Interviewer manual" and information provided directly by the National Institute for Statistics.

Table of contents:

A. General characteristics

Official name of the survey/data source:
Romania Integrated Household Survey (RIHS) / *Ancheta Integrata în Gospodarii*

Administrative Unit responsible for the survey:

National Institute of Statistics
Address: Libertatii Avenue, no.16, sect.5, Bucharest
E-MAIL : andreeav@insse.ro

The Romanian Integrated Household Survey (RIHS) started in 1994 as a World Bank Living Standards Measurement Study (LSMS) type survey providing data on household income and expenditures as well as the socio-economic characteristics of households. The survey was administered by the Romanian National Commission for Statistics (NCS) in cooperation with the Ministry of Labour and Social Protection and with the technical assistance of the World Bank. Today the National Institute of Statistics administers the survey.

The RIHS is meant as a permanent research upon a monthly sample of 3,000 households. It is the first large-scale nationally representative survey in Romania (the Romanian Family Budget Survey, conducted in 1989 and 1993, was a large scale data collection, but the sample used was not nationally representative). The RIHS collects detailed information on income and expenditure patterns of households, housing conditions, some aspects of health, fertility, education, and anthropometric outcomes for children for approximately 2,600 households each month.

The main aim of the survey is to obtain detailed data, structured by households categories, useful for: analyses of the most important aspects of living standard (demo-

social-economical characteristics of households and persons; incomes, expenditures and consumption; education; health status; working activities; labour force and unemployment; housing and endowment with durable goods etc.); the identification of disadvantaged categories of population; monitoring of transition effects on living standards; elaboration of a set of indicators for poverty measurement; creation of a statistical base for social protection programs.

The main users of the data are the Department of Consumer Price Index from the institute, ministries and research institutes.

B. Population, sampling size and sampling methods

Coverage

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.).

Sample size

The RIHS is meant as a permanent research upon a monthly sample of 3,000 households. The sample used for the 1997 dataset comprises 32,187 interviewed households (out of 36,072 sampled households) containing 92,334 individuals.

Sampling design

The sampling design was a two-stage procedure with a certain number of areas selected throughout the country in the first stage, and a fixed number of households selected in these areas in the second stage. The first stage sampling frame was a "master sample list" developed by the NCS with the aim of fulfilling all future needs of household and other surveys until the next census.

In the first stage it was selected "EMZOT master sample", 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), realised with the purpose of serving as samples basis for researches by testing within the dwellings, on an inter-census period. 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 network 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.

In the second stage it was selected dwellings from each research centre. Individual dwellings weren't selected directly. Groups of 3 dwellings were selected based on a randomly selection algorithm. The name of these groups is clusters and they are assimilated with *secondary selection units.* The algorithm of randomly selection it was based on calculation of selection interval and a random start for each research centre. From each research centre it was included, monthly, in the sample two clusters, 6 dwellings and 72 dwellings annually. For the sample dimensioning in the second stage (36072 dwellings) it was considered 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%.

C. Data collection and acquisition

Data collection period

All interviews were carried out between January 1997 and December 1997.

Investigation forms

The following forms will be filled in for each household in the lodging:
- the *investigation questionnaire*, including 19 sections to be filled in by the interviewer (see section E. below for details on its contents);
- the *household book*, including two chapters: *i)* the incomes and other money-inputs of the household members, which registers all cash returns and wage – payments in kind (originating from gratuities or price-cuts), both for the household as a whole and for each household member; *ii)* the household members' expenses and other payments, which registers all household ,embers' expenses for purchasing farm-food stuffs, non-food products, service payments, other payments during the reference month; the white leaf of the household book will register, as "other household notes", various information which the household members or the investigators make, and which are useful to the questionnaire registrations.

Survey organisation

Data for the RIHS were registered in two ways: first by filling out the household questionnaire according to answers provided by respondents to household interviews; and second, by having the respondents fill out a household income and expenses diary which recorded details of the household's incomes and expenses during the reference month. All the information entered in the expense diary was transcribed into relevant sections of the questionnaire by the interviewer so the each questionnaire contained a complete set of information for each household.

Responses were collected for all members who contributed fully or in part to the household budget (incomes and/or expenses). This includes household members present in the household, temporarily absent from the household, or absent for a longer period of time (more than 6 months). Responses were provided by the specific household member to whom the question pertained. If this was not possible, responses could be obtained from the household head or any other adult household member who could provide the information.

Considering the complexity of the questionnaire, and the various reference periods for which data were collected, information was collected by means of three compulsory visits to the household, along with supplementary visits, if necessary. Between the 1st of the reference month and the 10th of the subsequent month, the interviewer was to pay the compulsory visits as follows:
-   first visit: between the 1st and the 15th of the reference month;
-   second visit: between the 16th and the last day of the reference month;
-   third visit: between the 1st and the 10th of the subsequent month.
Supplementary visits were to be set up as required, with the agreement of the pertinent household members.

Survey implementation personnel

Overall supervision and coordination of the field work was done by the investigation manager and the head investigator who were in charge of organizational, administrative and managerial issues. The chief of computer operations was in charge of all computer related operations and the transmission of the data to NCS headquarters.

The responsibilities of the inspectors were to check, manage and occasionally assist the interviewers in implementing the survey in the field. They performed tasks such as assisting the interviewers in locating the households to be interviewed, checking for consistency between various sections of the questionnaire and checking for consistency between the questionnaire and the household expense diary, and helping with computer data entry.

The interviewers were responsible for carrying out the household interviews under the supervision of inspectors. They also checked the household members' entries in the household income and expense diary, sometimes physically recording the entries in the diary by the interview if the diary was left incomplete or if the respondents were incapacitated.

D. Definition of the survey units

Household

A household is a group of two or several people who usually live in the same lodging, are related, and keep the household in common, by sharing the incomes and expenses, totally

or in part. The household members work the land together, if necessary, raise the cattle, consume and capitalize together the goods they produce.

The household can also be constituted of a group of two or several people, childless or not, unrelated, but who consent to live and hold a common budget.

The people who do not belong to another household, but live a keep a common household, will be considered one-member households.

Household Members

Household members are considered the following:
- **present persons, temporarily absent ones or those left for other localities, on holidays, to visit their relatives, to rest homes, in the hospitals etc.;**
- **persons left for work, if they did not constitute another household or did not settle their domicile in other localities, still taking part in forming and using incomes;**
- **persons left for their military service or call-up;**
- **children in family placement;**
- **children in nurseries, kindergartens, camps, or supported by grandparents and relatives;**
- **institutionalized persons (houses for old people, houses for handicapped persons, monasteries) which are part of household, are absent in the reference period, but not absent more than a year;**
- **elderly people and other supported persons;**
- **persons arrested and convicted which are part of household, are absent in the reference period, but not absent more than a year**

The following people were not considered household members:
- the people hired for household work, the relatives who do not share the same household budget, the temporary guests to the house;
- the hosts or tenants, even if they temporarily (or for longer periods of time) live in the household and eat there.

Household Leader

The household leader is the person designated and accepted by the other household members; usually it is the husband. Accepting the household leader is the exclusive decision of the household members. If they hesitate to do so, the investigator may suggest other criteria, such as: economic status, income, age, personal authority, ownership of the property title or principle tenant.

Family Nucleus

The family nucleus is considered to consist of:
- the childless husband and wife;
- the husband and wife, parents of unmarried children;
- one of the parents who have unmarried children.

The family nucleus includes the unmarried children, even if they result from legal or illegal marriages, from a former marriage or from adoptions. The family nucleus also includes formerly married children who, during the reference month were widows or divorced and have returned to their parents, on condition that they do not support children from a former marriage.

A household can include one, two or more family nuclei, as well as no family nucleus at all (a household including only unrelated people).

E. Contents

The RIHS used a household questionnaire and a family diary to collect information. Participants in the survey were provided with diaries in which they recorded monthly cash flows (incomes, expenditures and savings).

The RIHS household questionnaire contained 19 sections, each of which covered a separate aspect of household activity. The various sections of the household questionnaire included:

Section	Title	Contents
0	Identification data	Research-centre code, investigation month, form number, lodging address and code, household number, interview result, visit dates, investigator's details
1	Household Roster	sex, age, relationship with the head of household, nationality, marital status, level of education, main occupational status, presence in household etc.
2	Present Occupational Status	occupational status, reason for inactivity,
3	Wage Activities	occupation, branch, employment modality, incomes from wages etc.
4	Non-Agricultural Non-Wage Self-Employment Activities	occupation, branch, incomes and expenditures
5	Pensioners, Beneficiaries of other Social Benefits	obtaining of social protection benefits (by benefit type each person receives)
6	Agricultural areas and administration type for the fields	agricultural surface, modality of land exploitation
7	A. Agricultural Capital Items B. Household durable goods	A. household endowment with different installations and machines for agriculture B. household endowment with main durable goods (by type, number and year of purchase)household endowment with main durable goods (by type, number and year of purchase)

8	Livestock	existence of animals and poultry stock and numerical evolution of this stock
9	A. Non-Household Members who Ate with Household Members B. Household members who Ate with other Households C. Meals served in public alimentation units	A. Number of days non-household members ate with household members B. Number of days household members ate with other households C. Meals served in public alimentation units
10	Food Expenses	food products balance sheet of household (quantities, by type of products, main sources of provenance and modalities to use)
11	Non-Food Expenses	purchases of non-food products (by type of products, quantities and values)
12	Expenditure on Services	expenditure for services payment (values, by services' type)
13	Incomes and Other Money Inputs	balance sheet of money income (by main providence sources of incomes, money balance, savings)
14	Expenses and Other Money Outputs	Balance sheet of expenditure (by expenditure categories, money balance, savings)
15	Incomes and Outputs of Cash in Reference Month	
16	Cash Savings in Reference Month	
17	Dwelling Information Characteristics	main dwelling of household (modality in which household is living in (owner, lodger, free of rent), dwelling's type, habitable surface, number of living rooms, technical characteristics, installations and utilities, household opinions about their dwelling and opportunity to move from it etc.)
18	Rest-Leave, Holiday Leave, Pleasure Holidays	
19	A. Health B. Usage of Iodate Salt	A. health status of household persons (by illness and disabilities type, treatment modalities, time and amounts of money needed for health care, treatment and recovery etc.) B. usage of iodate or other salt

F. Quality of data

Non-response

The overall non-response rate (i.e., percentage of sampled households not responding) for the year 1997 was 9.6%.

Data editing, validation, imputation

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 take 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..

There are automatic imputing procedures based on logical correlation's, so that there are no missing data.

Weighting

The grossing up of the results at the country level was done based on the coefficient assigned to the persons from the household comprised by the 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 extension coefficients.

Calculation of the basic weights **-** The basic weights of the dwellings (households) represents the first estimation of the extension coefficients. For their determination it was needed two 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:

  $$P_{1U} = \frac{n_{1U}}{N_{1U}}, \text{ where:}$$

  n_{1U} - number of dwellings from urban area included in EMZOT;
  N_{1U} - number of total dwellings from urban area;

  For rural area:

  $$P_{1R} = \frac{n_{1R}}{N_{1R}}, \text{ where:}$$

  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.

  $$P_{2k} = \frac{1}{I_k} = \frac{n_{2k}}{N_{2k}}, \text{ where:}$$

  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.
  $k = 1,\ldots\ldots,501$

- The calculation of the selection general probabilities of the dwellings in the research centre k, on urban, rural areas (PGS$_{kU}$, PGS$_{kR}$).

  For urban area:

$$PGS_{kU} = P_{1U} \; x \; P_{2k}$$

For rural area:

$$PGS_{kR} = P_{1R} \; x \; P_{2k}$$

- The calculation of the dwelling (household) basic weights in the research centre k, on urban, rural areas (PB_{kU}, PB_{kR}) .

  For urban area:

  $$PB_{kU} = \frac{1}{PGS_{kU}}$$

  For rural area:

  $$PB_{kR} = \frac{1}{PGS_{kR}}$$

The basic weight (PB_k) is the first estimation of the extension coefficient. This basic weight is a weight for all the dwelling in the research centre k, but it is use also to the households living in these dwellings and implicitly to all the persons from the households.

The non-response adjustment - To cover the percentage of the households the refuse 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 next variable: region, county, area and occupational status of the household head.

- Adjustment rate with non- response for each cell. In this way, the households were grouped using these cells and the non-responses were treated separately on each cell.

$$RNR = \frac{\sum PB \; ES}{\sum PB \; RASP \; ES} \text{, where:}$$

RNR – adjustment rate with non-response;
$\sum$ PB ES - the sum of the basic weights of the households from the sample;

$\sum$ 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:

$$P_{AJ} = PB \cdot RNR$$

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 was used like auxiliary information, demographic variables (sex, groups of age) and localisation variables (area of residence, region) disposable from demographic sources (legal population at first of July). 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 are made cells of persons crossing the variables: area/sex/groups of age, from the sample and from the total population. For each region, in each cell: area/sex/groups of age is multiplying the weight obtained in the steps before, with a coefficient of final adjustment (C_r):

$$C_r = \frac{Total\ number\ of\ persons\ by\ region\,/\,area\,/\,sex\,/\,group\ of\ age}{Number\ of\ weighted\ persons\ from\ sample\ on\ region\,/\,area\,/\,sex\,/\,group\ of\ age}$$

At the end of this step, the final adjustment coefficient assigned to each person from household is:

$$COEF_{ext} = P_{AJ} \cdot C_r = PB \cdot RNR \cdot C_r$$

General measurement problems

There are three broad measurement problems in the RIHS.

1. The survey collects information on income and consumption of households using a diary method that forces household income and expenditure to strictly match with each other. The NCS asks individuals to keep a monthly diary recording household cash flows. Interviewers check to see whether incomes (cash in) strictly match the sum of expenditures (cash out) and net change in savings for that month. This means that consumption is not measured independently of income and is subject to the same measurement errors as income. Essentially, if individuals under-report income, consumption will also be under-reported.

2. It is difficult to construct income from the data. Very few households report self-employment income and most households fo not market their produce. Therefore, measurement of self-employment income is unreliable and annual agricultural income is extremely difficult to estimate. Although the survey provides a detailed breakdown of the components of agricultural production and associated input costs, it does not provide information on labour costs or the timing of inputs, making the estimation of profits extremely difficult. Furthermore, since few households market produce, the information on market prices is very thin. The survey also does not include a community price questionnaire. Prices can only be proxied by unit values of purchased food goods. There is an additional problem; the survey provides quantity of food goods purchased and the value of goods purchased both from the state and the private sector, Thus, prices used in the computation of consumption are quantity weighted average of state and private prices (unit values).

3. The RIHS uses a cash flow concept to construct aggregate household consumption and income. Although wages and other fixed incomes are defined as in other data sets, agricultural and self-employment income are revenues from sales, not profits. Similarly, aggregate consumption data provided by the RIHS was based on purchases of food, non-food goods, services and durables. However, food consumption is not solely comprised of purchases (some of which may actually be stored). Households can consume from home production, by depleting existing stocks of goods, or through gifts/transfers of food from friends/relatives/employers.

G. Uses of the survey

Publications

Each year two publications are published on the basis of survey results: "Population quality of life aspects" and "Income, expenditures and consumption". Also data from this

survey was disseminated in the: Statistical yearbook, publications on poverty, UNICEF publications etc.

Poverty

Table "Alternative estimations of the overall poverty rate", from Study on the Social Protection Systems in the 13 Applicant Countries: Romania – Country Study", January 2003, Study financed by the European Commission – Employment and Social Affairs DG

	Prate CAS	Prate EUR	Prate RGE	Prate RGWE	WB Est	Tesliuc Es	Prate RGWO	WBA Est	Prate CASO
1989			*94.5*		3.9			17.7	
1992			*27.7*	*20.7*	70.9		*39*		
1993			*39.8*	*46.9*			*40.6*		
1994			*40.2*	*46.2*	27.5		*42.1*		28
1995	25.2		*35.4*	*58.1*		25.3	*43*		25
1996	23.3		*40.2*	*52.5*		19.9	*44.6*		20
1997	27.7		*34.5*	*65.9*		30.1	*47.6*		30
1998	27.3		*36*	*54.7*	44.5	33.8	*46.3*		34
1999	26.6	39	*39.4*	*47.6*		41.2	*47.1*		41
2000	30.6		*41.7*	*35.9*			*48.8*		44
2001	29.6		*38*	*39.3*			*39.3*		
2002	23.4		*39.5*	*31.6*			*31.6*		

Figures in italics stand for the independent estimates of the author of this table (C. Ghinararu), computed after performing several regressions between overall poverty rate as dependent variable and several independent variables (notably GDP% growth and the share of wages in the total income of individuals).

1) PrateCAS-stands for the poverty rate as calculated by the National Anti-Poverty and Social Inclusion promotion Commission, after a methodology which was elaborated with the assistance of the World Bank, the UNDP and several Romanian think tanks (research institutes, universities). It stands today as the official poverty rate assessment;
2) PrateEUR – stands for the „at risk of poverty rate (poverty rate) before transfers, calculated by EUROSTAT, after its own methodology;
3) PrateRGE – stands for the reconstructed EUROSTAT series for poverty rate (at risk of poverty rate, before transfers), as computed by the author of this paragraph after performing a regression using GDP% growth as independent variable;
4) PrateRGWE – stands for the reconstructed poverty rate series (at risk of poverty rate before transfers) computed by the author of this paragraph after performing a regression using wage share as independent variable. At its turn, wage share has been computed after performing a regression using GDP% growth as independent variable;
5) WBEst – stand for World Bank estimates as published in the World Bank Development reports published throughout the period as well as in the dedicated World Bank study „Making Transition work for everyone. Poverty and inequality in Europe and Central Asia", published in 2000;
6) TesliucEs – stand for Emil Tesliuc estimates for the World Bank; Emil Tesliuc is a Romanian economist currently with the World Bank;
7) PraRGWO – stands for the reconstructed series of the poverty rate (at risk of poverty rate before transfers) as computed by the author of this paragraph after performing a regression using as independent variable the official series for the share of the wages in the total income of the individuals;
8) WBAEst – stands for another World Bank estimate published in one of the World Banks reports of the period, that assesses poverty rates at 1985 PPP rates;

9) Prate CASO – stands for series computed by the National Anti-Poverty and Social Inclusion Promotion Commission after an earlier methodology that has been replaced by the one elaborated with the assistance of the World Bank (this methodology is no longer in use).