Data and Methodology[1]

This document summarizes the main features of the survey design, data and methodology used in the Bulgaria 2001 Integrated Household Survey.

Survey design and sample

The 2001 Bulgaria Integrated Household Survey (BIHS01) was conducted by BBSS Gallup Int'l. under the supervision of the World Bank. The BIHS01 is the third such data collection effort by Gallup/World Bank in Bulgaria since 1995. The first BIHS was conducted in 1995 on a sample of approximately 2,500 households. The second round was conducted in 1997 on the 1995 sample. A total of approximately 2,000 were re-interviewed in 1997. Because of the expected excessive level of attrition due to the large time lag from the last survey and the massive internal and external migration since 1997, for the purpose of this survey it was decided to draw a new cross-section of households. Using the same stratified two-stage cluster design adopted in 1995, a similar nationally-representative sample was drawn by the National Statistical Institute (NSI) from the pre-census listing of the 2001 Population Census.[2]

Although special instructions and particular care was to be devoted to reducing refusals to a minimum, a decision had to be taken on a new household replacement rule, as the one adopted in 1995 was deemed unfeasible due to time constraints. As in 1995, the original sampling plan called for the selection of five households in each of 500 randomly selected census clusters. In 2001, six households per cluster were provided by NSI to Gallup and the sixth household was used to replace households in the original sample in cases of refusal or absence. Each field substitution had to be verified by the team leader and approved by the field supervisor. A total of 2,500 households were finally interviewed. In addition, 133 Roma households were oversampled to allow more significant statistical comparisons of the group in some of the analyses. Detailed rules for the selection of the oversample were given to the enumerators and each selection was verified by the team leader.

As the main objective of the survey was to provide comparable poverty figures with the previous studies, the questionnaire used is virtually identical to the one used in the previous surveys and – when changes were introduced – particular attention was paid to maintain consistency with the previous questionnaires.

Being a multi-purpose survey, the BIHS01 questionnaire follows the structure of a typical Living Standard Measurement Survey (LSMS). The survey collected exhaustive information for the estimation of a consumption aggregate. This includes food and non-food consumption expenditures as well as data for the imputation of housing rental value and the user value of durable goods. The questionnaire also contains comprehensive information for the estimation of income by source, as well as quite extensive information on health, education and the labor market.

[1] This document was written by Gero Carletto (DECRG) and Tomoko Fujii (consultant) as part of the poverty assessment for Bulgaria, World Bank Report No. 24516.

[2] Information on the 1995 and 1997 Bulgarian Integrated Household Surveys can be found on the LSMS web site: http://www.worldbank.org/lsms/lsmshome.html

Field workers' training was conducted by Gallup, with World Bank supervision, in Sofia and Varna in March 2001. The survey was fielded starting in early April 2001 and completed at the end of May 2001

The consumption aggregate

Consumption as defined for this survey is the monetary value of all food and non-food goods and services consumed by households. It includes all purchases as well as the value of home-produced goods, and the value of goods and services received in kind. Non-food goods cover clothing, cleaning, transport, utilities, health, entertainment, education and housing utilities.

To define aggregate household consumption, four issues had to be addressed: consumption of home-produced food, valuation of health care and education, treatment of consumer durables and the treatment of owner occupied rent.[3]

Home-produced food is a very important source of consumption in Bulgaria. Excluding home-produced food would bias consumption downward. The consumption of milk was found to be significantly higher for those households with cheese or yogurt production. Therefore, an adjustment was made to accommodate cheese or yogurt production to reflect the pure consumption of milk. To value home-produced food, the actual current price reported by the respondent is used. When such a price is not available, the regional or national median price was used instead.

Out-of-pocket health care expenses and insurance are included in non-food expenditures. The health care items asked in the survey were dentist, doctor, hospital, medicine, optical, skin care and other health related expenses. Inclusion of health expenditures increases the consumption aggregate and hence the welfare measure of households with sick people. Exclusion of health expenditures would result in the same consumption aggregate level for households with different abilities to pay for health care, if they consume the same amount of everything else.

Durables are not purchased frequently. Including them in the estimate of consumption would exaggerate the permanent purchasing ability of the household and lead to misclassification of its poverty status. Theoretically, the flow of the service should be included. In spite of the fact that information for the estimation of user value was available in the 2001 survey, to ensure comparability with 1995 estimates we did not include durables in the consumption aggregate.

Housing is not included in the computation of consumption aggregate. Housing is integral part of one's living standards and often accounts for a non-negligible part of expenditure. However, we decided not to include the rent due to the following reasons: (a) the majority of people own their house and the rental market of housing is very small in Bulgaria[4]; and (b) rents were not included

[3] All programs used to calculate the consumption aggregate are included in the documentation for the survey.

[4] Only 6% of the households rent houses, with a mean rental of 46.5 BGN and standard deviation equal to 46.2 BGN. Furthermore, more than half of renters are renting from the State. Rents for public housing (mean 30.1, s.d. 29.8, obs. 84) are substantially lower than private rents (mean 67.3, s.d. 54.5, obs 66). Those factors make it virtually impossible to compute reliable imputed rents.

in the consumption aggregate in 1995 or 1997. Therefore, to maintain consistency with previous estimates, it was deemed preferable not to include housing this time as well.

Regional price adjustment

Substantial spatial price variation was observed during the administration of the BIHS. Failure to account for the price differences across regions, and between urban and rural areas would result in misleading welfare measures. For example, given the nominal expenditure level, households would be better off with lower prices because of the higher purchasing power.

A total of 18 price indexes were estimated, one for urban and rural for each of the 9 regions. Each index was derived by aggregating a food and a non-food price index. For the calculation of the food price index, price data on all major food items were available from the BIHS. The bottom 40 percent was used as the reference population, as we deemed using prices paid by the poorest as more relevant for the purpose.

To ensure that enough observations were available for each of the food items, 7 groups were formed out of the 78 food items and average prices were computed for each group as the ratio of total expenditure to the total quantity. The food price index for each of the 18 locations was computed as a weighted average of the group indexes, where the weights were the budget shares of each group out of total food expenditures for the reference population.

With regard to non-food items, a similar procedure was used. However, only the prices for alcohol, tobacco, cleaning, personal items and gas and oil were available. The assumption was made that the non-food price was equal to the non-food prices calculated by those five items. In other words, the shares of expenditures on those items in the basket are scaled up to cover the entire non-food expenditure in the basket. Finally, the two indexes for food and non-food were aggregated by means of a weighted average.

The results of the calculation are reported below:

	Urban	**Rural**
Sofia	1.08	1.03
Bourgas	0.99	1.01
Varna	1.01	1.03
Lovech	1.02	0.96
Montana	0.98	0.88
Plovdiv	1.08	1.06
Russe	1.00	0.91
Sofia Region	1.00	1.05
Haskovo	0.99	0.98

Seasonality adjustment

Survey data, including the BIHS, are typically carried out at a given point in time. This implies that, without seasonal adjustment, the consumption aggregate based on the survey data reflects the preference of the household which may be affected by seasonal availability of goods and services in the market and the seasonal change in needs of certain items. Therefore, to the extent that such seasonal patterns are significant, the mean consumption expenditures of food and non-food items exhibiting seasonality will not yield a representative picture of the average consumption expenditures of households throughout the whole year.

The seasonal adjustment applied to the data is based on work done originally for the 1995 survey by Skoufias. He used the 1994 Household Budget Survey (HBS) collected by the Central Statistical Office of Bulgaria. He divided food items into five groups and nonfood items into eighteen groups. He also divided the population by income group (top 20%, middle 60% and bottom 20%) and by location sector (Sofia, Other urban and Rural areas). Therefore, he had 9 sub-samples for each of 25 commodities. He used the household-specific dummy variable approach that accounts for the influence of time-invariant characteristics, including observable and unobservable characteristics. Coefficients associated with the dummy variables for each month and others were estimated by an OLS regression for each sub-sample and an F-test of the joint hypothesis that there are no significant differences in the average per capita consumption across months were tested. The correction factor for each month was, then, calculated as the ratio of mean per capita consumption of that month over the monthly mean per capita consumption averaged through 1994.

At the time the BIHS01 was first analyzed no data were available from a more recent HBS. We assumed that the seasonal consumption pattern had not changed significantly since 1994 and used the same correction factors based on the 1994 HBS.[5] As the dataset of the 2000 HBS became available, it became possible to re-estimate the aggregates based on the newly computed seasonal adjustment factors. These re-adjusted figures are provided as well as the originally created files. To do comparisons to the 1995 and/or 1997 data, the 1994 seasonal adjustment factors must be used.

Equivalence scales

Using total household consumption would be a misleading measure of the welfare level of its members as it does not take into account differences in household size and composition. The per capita consumption measure was used to allow for such differences. However, there are shortcomings in the per capita consumption measure. First, per capita consumption does not allow for differences in needs. Children are assumed to have the same needs as adults. Second, it does not allow for the economies of scale in consumption. Two can often live as cheaply as one.

Introducing an equivalence scale assumption is a way of adjusting for differences in needs in households of different size and composition. However, such adjustment tends to be subjective and there is no accepted equivalence scale in Bulgaria. Also, per capita measures were used in the studies of the previous surveys. Per capita expenditure is calculated here to avoid introducing subjective equivalence scales and to allow comparability with previous BIHS figures. Thus, care should be taken in interpreting demographic profiles as they are generally very sensitive to assumption made on equivalence scales and economies of scale in consumption. Let it suffice to note for now that inclusion of equivalence scale and economies of scale generally tend to weaken the correlation between poverty and family size. This is because large

[5] Although the majority of interviews were conducted in the six-week period between mid-April and end-May, we adjusted for seasonality using the May factors, due to non availability of the April factors.

families with many children may have low per capita income, but may not have as many adult-equivalent people and may not be as poor due to substantial economies of scale in consumption.

Choice of Poverty Line

To assess the welfare status of each individual, a common yardstick is required to measure the estimated per capita consumption level. The poverty line sets the threshold below which an individual will be considered poor. Such threshold can be set in absolute terms, generally associated with the cost of a basket containing a set of food providing a minimum nutritional requirements plus some basic necessities, or in relative terms, defined in relation to some parameter (e.g. the mean or median) of the sample distribution.

Bulgaria lacks an official poverty line. Many poverty lines are available but none enjoys either unanimous support or official status in the country. For the sake of comparability with the 1995 and 1997 estimations, two relative lines are chosen for this study at one half (extreme poverty) and two thirds (poverty) of median consumption in 1997, deflated at March 2001 prices based on the Consumer Price Index. The deflated 1997 value corresponds to a higher poverty line of BGN 61.5 and a lower line of BGN 46.1. However, to put these chosen lines in due perspective, we compare them with some other lines used in Bulgaria (see Box 1).

BOX 1. COMPARING POVERTY LINES

The, the choice of relative poverty lines anchored to the 1997 consumption levels was dictated by the necessity to provide comparable figures with the only published report from the previous BIHS.[1] However, it is useful to put the chosen poverty line in perspective, to facilitate comparison with other studies and across countries.

Although no official poverty line exists in Bulgaria, a number of lines are currently used for different purposes. The Guaranteed Minimum Income and Minimum Social Pension in 2000 were set at BGN 37.4 and BGN 40, respectively. These levels are somewhat comparable to our *lower* poverty line, set at BGN 46.1. It must be noted that the benefit eligibility lines are driven by budgetary considerations and by no means reflect suitable consumption needs. In the table below, we report several poverty lines and compare the corresponding poverty rates. The *higher* relative poverty line used in this report is comparable with the Cost of Basic Needs line for 1995, deflated at 2001 prices, which assumed a food share of 0.68 (del Ninno, 1996).

TABLE B1.1. Comparing poverty lines

	Level (2001 BGN)
Guaranteed Minimum Income(*)	37.4
Minimum Social pension (**)	43
Lower poverty line	46.1
$2.15 PPP line	47.9
Cost of Basic Needs (***)	57.0
Higher poverty line	61.5
Subsistence minimum (***)	86.0
$4.30 PPP line	95.8

(*) latest available figure is 2000. Based on past trends, assumed that no adjustment to the level has been made in 2001.
(**) latest available figure is BGN 40 at 2000. Assuming a CPI deflator of 1.066, the 2001 level is assumed equal to BGN 43.
(***) Based on deflated 1995 estimated cost of food basket to consume 2,100 kcal using bottom consumption quintile. Cost of Basic needs assumes a 68% food share, while in the subsistence minimum the share is set at 50%.

[1] Bulgaria: Poverty during the Transition, World Bank, Report No. 18411, 1999.

Appendix 1. Guidelines on the Use of the Data Set

1. In the data as originally key entered “refuse to answer” and “don’t know” were coded as 97 and 98 respectively. These have been recoded to –1 and –2.

2. hhnumber==925 is dropped in the cleaned version of the data due to the quality of the data in food section, and possibly other sections.

3. idcode==. | idcode==0 is dropped in the cleaned version of the data because such data contains no significant information. Only relevant in r1.dta (3 obs), r2.dta (1 obs) and r6_1.dta (4 obs).

4. Variable names were converted to a more "meaningful form[6]" from the r*_q* form in the data cleaning process. The r*_q* form makes it easier to understand the link between the questionnaire and the variable in the dataset whereas a “meaningful form” makes it easier for the programmers to write and maintain programs. To see the correspondence between the question, users are referred to CodeMapExt.xls

5. To merge data sets together, “hhnumber” is used to merge at the household level and “idcode” is used as the personal identification number.

6. BGN refer to New Lev. In July 1999, the Lev was redenominated so that 1000 Lev (BGL) now are equal to 1 BGN.

[6] For example, for question 19 of the household roster, “gender of the main respondent”, the variable name is “r0_q19” in the original key entry of the data and it is changed to “gender” in cleaned version of the data.

Appendix 2. Problems Encountered During the Processing of the Data

1. Things need to be noted.
1.1 District, location type file must be updated.
THIS HAS DONE ALREADY

(TERMINOLOGY)
District: Biggest geographical division. Numbered 1-9
Region: Biggest geographical division in the data set. Numbered 1-28
Cluster: Geographical division by which sampling is based. Numbered 1-500
Location type: Rural/urban (1 or 2)

We may wish to collect basic demographic information on each cluster
for future analysis. (Male/Female/Under 18 population, Number of Households, etc).

1.2 Seasonal Price Adjustment
Seasonal price adjustment is also necessary. We have relevant documents by Emmanuel Skoufias. We still need some data to compute seasonal adjustment based on the observation this year. We are using adjustw file from 1995. This has an implication that the seasonal consumption pattern has not changed since then.

1.3 No record for R3 and R10 for those with person id >10
THIS ISSUE CANNOT BE RESOLVED AT THIS STAGE.
(According to Radosveta), because of the questionnair design, it was not possible to enter the data for persons with personal id number greater than 10.

1.4 Some low quality observations
LOW QUALITY DATA IS ELIMINATED. Household 925 is the only observation eliminated. This is done when you run the do files in DATACHECK folder. Missing idcode and idcode==0 are also eliminated for data sets which idcode is available.

1.5 Known data problems
*r0
--hhnumber 466 & 495
There are two household heads in those two households. It seems that idcode=1 is the household head.
--hhnumber 786
There is no household head. By the spouse relationship code, it seems that idcode=1 is the household head.

*r6_3, q16&q17m, q17y uid 1610, rch1015204
We have one observation for q17m and q17y (Nov 2000). But this household answers that it received no subsidized vacation in q16.

*r6_6_1
There is one record for which crop ==0 (uid 1520). This observation can be ignored.

*r6_6_2
There are three records for which crop==. (uid 1104, 1753, 1866)

*r8_4, q8m, q8y
uid 2348, rch 1729006
uid 2505, rch 1420409
They answer that they have received the benefits after May 2001. (Dec 2001 and Jul 2001, respectively).

*r8_9, q2y q2m q3y q3m q5
bcode is the code for the type of benefit and takes a value between 1 and 7. There are some households that say they stopped receiving a certain type of benefits before year 2000, but answer in question 5 that the total amount they received is positive (don't know for 2253-4)

uid	rch	bcode
3	2137001	2
359	2545302	6 *
1050	101914	4
1970	611703	2
2056	1728704	4
2253	1727609	3 *
2253	1727609	4 *

For records with *, the household starts receving the payment after it stopped receiving the payment.

*r8_10 code
The value label for "8" was not given. There are seven instances of this. The uid's are 140, 381, 449, 491, 1022, 1518 and 1861.

*r9_2 q3
There are three records in which the answer to q3 is zero. The uid's are 1521 2006 and 2016. Those records seem to contain no substantial information.

1.6. Price Adjustment according to inflation

(old:priceadj.do)
(new:Deflator.do)
Adjustment for inflation is done by using the deflators calculated in Deflator.do, which is used in LivestockAggregate.do. The CPI for April and May 2001 is not yet available. Those CPIs must be replaced.

2. Potential methodological concerns

2.1 Treatment of rent

(old:CLCEXP04.DO)
(new:HOUSEAGGREGATE.DO)

In 1995 program, rent is treated in the following way: Use if information on rent is available. Or else, use the national median 82.2. I don't know where this number is taken from. For 2001, I used observed median.

2.2 Treatment of outliers for utilities and education

(old:CLCEXP04.DO & CLCEXP05.DO)
(new:HOUSEAGGREGATE.DO & EDUCATIONAGGREGATE.DO)
There are a number of hard coded numbers that determines the border between outliers and normal values. I checked the distribution. Most of the cases, only several percent of the observations are eliminated. I instead set an upper limit of the value, which is mean plus five times standard deviation. Although this criteria is arbitrary, assuming normal distribution (with wishful thinking), it is a very very unlikely observation. There are several percent (often less than 1%) of observations treated as an outlier in both 1995 and 2001. A major exception is d_yr_tut (q27 in 1995 and q32 in 2001 in education section). The detail is explained in the next section.

2.3 Treatment of special training/tutoring

(old:CLCEXP05.DO)
(new:EDUCATIONAGGREGATE.DO)
The quesiton asks "How much was usually paid per month "..." special training/turtoring during last school year?" Obviously, the amount must be a per month figure, but the program divides this by 12. Also, although there are only 99 observations for this, 23 are considered as an outlier due to the "hard coding". My criterion (5 sd + mean) gives 12 outliers out of 119 observations. In 2001 also, the amount of training/tutoring is still divided by 12, but this is a clear mistake.

2.4 Treatment of price and quantity in food sector

(old:EXP11C.DO and CLCEXP01.DO)
(new:FoodAggregate.Do)
I shall only discuss prices here. Quantity is treated in a similar manner. See the do file for details.

If a price is three times higher than the mean AND higher than twice standard deviation plus the mean, then that price is "impossible". They treat that price as missing. Including such price, the missing prices are dealt wtih in the following way:
1) Use the regional median price if it is available
2) Use the location sector (Urban/Rural) price if it is available
3) Use the national median price

For "truly" missing prices (or don't know/ refusal to answer), those rules are, arguably, all right. However, for the treatment of "impossible" prices, I have to say this is a seriously flawed treatment. Suppose first that three times mean is higher than two times standard deviation plus mean. Let us assume the regional median is $5 and three times mean is $16. Now, suppose person A is paying $15.90 and person B is $16.10. Then, according to the treatment used in 1995, $16.10 is deemed to be impossible. Thus it must be replaced by $5, substantially understating the consumption of person B. In other words, if someone is paying a much much higher price than the average (perhaps in the pursuit of quality of the goods), then such person is supposed to be spending the regional median price.

Also, assuming normal distribution, about 2.5% of the people are paying more than twice standard deviation plus mean. If the standard deviation is greater than mean, then the prices they are facing are qualified for being "impossible" prices. This is also problematic. The real problem with outliers, I think, is not that they exist but that one household can affect excessively statistical figures taken over the entire sample.

Potential remedy would be to set an upper bound or squeeze the too high prices into a certain interval. Of course, arbitrariness is unavoidable, but we can definitely avoid making a mistake of concluding that a household is in the middle class even if that household is paying 1000 times as much price as the median price for everything.

2.5 Treatment of own production

(old:EXP11C.DO, EXP11D.DO and CLCEXP01.DO)
(new:FoodAggregate.Do)
Consumption of milk is treated in a special manner. Milk is used for the production of yogurt and cheese. This is fine, but how milk was chosen for a special treatment is not clear. In comments of EXP11C.do, they claim that they have checed grapes and wine. Also in CLCEXP01.do, they claim, "After careful checking, we concludd that the only serious over-reporting could occur only in the case of milk and milk products." However, I could not find any trace of the analysis that would lead to this conclusion.

One of the ways to treat this is two employ the two-normal sample model. It is possible to divide the sample into groups with own production and without own production, and then carry out a t-test with the null hypothesis being U0 = U1, where U0 is the mean for the group with own production and U1 is the mean for the group without own production. This obviously requires some more programming, but potentially it is a topic of interest.

2.6 Treatment of total consumption

(old:CLCEXPPd.Do, CLCEXPPS.DO, MKPOORD.Do, MKPOOR.Do)
(new:PercapitaConsumption.Do, AdjustedPerCapitaConsumption.Do, PovertyProfile.Do)
When the consumption aggregate deciles (for both household aggregate and per capita) are created, those household that are considered as outliers are not included. 100 and 20000 are the hard coded lower and upper bound for normal values (nonoutliers).

In the old program mentioned above, there are a number of lines that go like:

```
gen decile = group(10) if pcexp>100 & pcexp<20000
```

Since pcexp is created by removing outliers for each item of household consumption, I am not sure if this procedure is justifiable, let alone necessary. For now, the "if" clause is removed. Only top and bottom 1% or so are removed in this procedure, and the poverty line should not be affected that much even if it is affected at all. Please see the next section as well.

2.7 Derivation of poverty line
(old: Mkpoor.do, Mkpoord.do)
(new: PovertyProfile.do)
Poverty line is hard coded. There is a comment that says it was derived by 2463 * 1.6 * 0.67 = 2637 (Well, this calculation itself is wrong). The derivation of poverty line will be discussed

2.8 Making CropAggregate
(old:clcexp1a.do)
(new:CropAggregate.do)
Prices for crops for which no observed price is avialable are determined exogenously. They are hard coded in the program and the source
of the data is not specified. Also, there are several hard coded upper bound above which the data are considered outliers. Due to the unavialability of the log file, it is at this stage impossible to see how many observations were excluded. (We can reproduce it by running the do file using available dta files, but this may be different from the original results.) I just commented out this part of the program. In Step 6, they have the following line.

```
replace kg_sold=50 if kg_harv~=. & kg_sold ==.
```

I don't know where this number 50 comes from. I define "sales ratio" and assume that the proportion of the quantity going to the market is equal to the ratio of median quantity old over median harvested quantity.
Also, in Step 6, there is a line like:

```
        replace cons=. if cons<5;
```

I do not know where this number 5 comes from.

2.9 Making Livestock Aggregate
(old:clcexp1b.do, clcexp1c.do)
(new:LivestockAggregate.do)
Firstly, clcexp1b.do uses priceadj.dta to allow for the seasonal price adjustment. For this purpose I need the consumer price indices. At this time, I have created a dummy data set, which needs to be replaced. Since there are a number of "don't know' and 'refusal to answer' within a very few number of observations

2.10 MainJob & SecondJob (Section 6.2 & 6.3)
(old:clcinc2.do, clcinc3.do)
(new:MainJobAggregate.do, SecondJobAggregate.do)
The threshold for the number of hours worked in one week (q6) to determine outliers seems too low (70hours). The way outliers are eliminated (especially for mj_hrs, mj_mths and mj_gros) is also weired. Please compare those two programs for details. I restricted the outlier elimination only if the number of weeks worked is less than 4 weeks. (i.e. if the number of hours worked is greater than 70 and the number of weeks worked is less than 4, then it is replaced by 40. Also, by the survey design, the maximum number of hours worked is less than 100. We allow missing values for this.

2.11 SelfEmployment (Section 6.4)

(old:clcinc4.do)

(new:SelfEmploymentAggregate.do)

They drop the observations without self_csh or self_knd. There are only 6 observations (about 3%) to be dropped, but the sample size is not very big.

2.12 Remittances received by hoseuholds (Section 7.1)

(old:clcinc8)

(new:RemittancesReceivedAggregate.do)

Question 9 and question 11 in 1995 do not exist the 2001 questionnaire. As a result, q_cloth and q_value are used instead of value and tot_val.

2.13 Cash and inkind benefit(Section8.8 in 1995 Section8.7 in 2001)

(old:clcinc7)

(new:IndividualInkindBenefitsAggregate.do)

Question9 in 1995 is asked in a different way from that in 2001. I treat the answers in the same way.

2.14 Cash and inkind household social benefits(Section8.9)

(old:clcinc7)

(new:HouseholdBenefitsAggregate.do)

We do not have Question6 in 1995 questionnaire(b_costbe).

2.15 Other forms of revenue-debt (Section8.10)

(old:clcinc7)

(new:OtherRevenueDebtAggregate.do)

We do not have Question4 in 1995 quesitonnaire(b_costbe). I use amount_yr instead.

2.16 Real Estate Assets (Section9.2)

(old:clcinc8)

(new:RealEstateAggregate.do)

We do not have Question12 in 1995 questionnaire(p_val_re). I use p_intere instead.

(*) A hard coded number refers to such a number that is not derived from anything in the data and is in this case just a constant placed in the program. For example,if you have a code like

```
gen upperbound = 1234
drop if exptot > upperbound
```

then 1234 is a hard coded number. If you have a code like

```
egen upperbound = egen(exptot)
replace upperbound = upperbound * 10
drop if exptot > upperbound
```

1234 is not a hard coded number.

Path	DoFile	In	Out	Description
CONSUMPTION				
	AdjustedPercapitaConsumption.do	adjustedtotalaggregate.dta householdsize.dta	adjustedpercapitaconsumption.dta	Computing pc consumption seasonally & regionally adjusted
	AdjustedTotalAggregate.do	totalaggregate.dta districtpriceindex.dta	adjustedtotalaggregate.dta	make regional adjustment based on seasonally adjusted hh consumption
	DistrictPriceIndex.do	percapitaconsumption foodtemporary.dta strata.dta percapitaconsumption.dta nonfoodtemporary.dta strata.dta districtpriceindextemp.dta	districtpriceindextemp.dta districtpriceindex.dta	make price index for each district
	EducationAggregate.do	r3.dta strata.dta	educationaggregate.dta	aggregate educational expenditure
	FoodAggregate.do	r5_1.dta strata.dta householdsize.dta gitemf.dta	foodtemporary.dta foodaggregate.dta	aggregate food expenditure
	Gitemf.do		gitemf.dta	define group items for food
	HouseAggregate.do	r4.dta strata.dta	houseaggregate.dta	aggregate house related(rent/utilities) expenditure
	HouseholdSize.do	r1.dta	householdsize.dta	calculate hsouehold size and adult equivalence
	NonfoodAggregate.do	r5_2.dta strata.dta	nonfoodtemporary.dta nonfoodaggregate.dta	aggregate non-food expenditure
	PercapitaConsumption.do	totalaggregate.dta householdsize.dta strata.dta	percapitaconsumption.dta	aggregate per capita consumption
	SampleFrame.do	sample.txt	sampleframe.dta	define sample frame
	SeasonalAdjustmentFactor.do	seasonaladjustmentfactor.txt	seasonaladjustmentfactor.dta	import seasonal adjustment factor and convert it into a usable form
	Strata.do	r0.dta r11.dta sampleframe.dta	strata.dta	make a stratification file (collect several attributes to household in one place)
	TotalAggregate.do	foodaggregate.dta nonfoodaggregate.dta houseaggregate.dta educationaggregate.dta householdsize.dta strata.dta seasonaladjustmentfactor.dta	totalaggregate.dta	total nominal (not adjusted) and seasonally adjusted consumption are calculated
READ				
	r0.do	r0.dct	r0.dta	Convert ascii file to stata file
	r1.do	r1.dct	r1.dta	Convert ascii file to stata file
	r10.do	r10.dct	r10.dta	Convert ascii file to stata file
	r11.do	r11.dct	r11.dta	Convert ascii file to stata file
	r2.do	r2.dct	r2.dta	Convert ascii file to stata file
	r3.do	r3.dct	r3.dta	Convert ascii file to stata file
	r4.do	r4.dct	r4.dta	Convert ascii file to stata file
	r5_1.do	r5_1.dct	r5_1.dta	Convert ascii file to stata file
	r5_2.do	r5_2.dct	r5_2.dta	Convert ascii file to stata file
	r6_1.do	r6_1.dct	r6_1.dta	Convert ascii file to stata file
	r6_2.do	r6_2.dct	r6_2.dta	Convert ascii file to stata file
	r6_3.do	r6_3.dct	r6_3.dta	Convert ascii file to stata file
	r6_4_1.do	r6_4_1.dct	r6_4_1.dta	Convert ascii file to stata file
	r6_4_2.do	r6_4_2.dct	r6_4_2.dta	Convert ascii file to stata file

	r6_4_3.do	r6_4_3.dct	r6_4_3.dta	Convert ascii file to stata file
	r6_5.do	r6_5.dct	r6_5.dta	Convert ascii file to stata file
	r6_6_1.do	r6_6_1.dct	r6_6_1.dta	Convert ascii file to stata file
	r6_6_2.do	r6_6_2.dct	r6_6_2.dta	Convert ascii file to stata file
	r6_7.do	r6_7.dct	r6_7.dta	Convert ascii file to stata file
	r6_8.do	r6_8.dct	r6_8.dta	Convert ascii file to stata file
	r6_9.do	r6_9.dct	r6_9.dta	Convert ascii file to stata file
	r7_1.do	r7_1.dct	r7_1.dta	Convert ascii file to stata file
	r7_2.do	r7_2.dct	r7_2.dta	Convert ascii file to stata file
	r8_1.do	r8_1.dct	r8_1.dta	Convert ascii file to stata file
	r8_10.do	r8_10.dct	r8_10.dta	Convert ascii file to stata file
	r8_2.do	r8_2.dct	r8_2.dta	Convert ascii file to stata file
	r8_3.do	r8_3.dct	r8_3.dta	Convert ascii file to stata file
	r8_4.do	r8_4.dct	r8_4.dta	Convert ascii file to stata file
	r8_5.do	r8_5.dct	r8_5.dta	Convert ascii file to stata file
	r8_6.do	r8_6.dct	r8_6.dta	Convert ascii file to stata file
	r8_7.do	r8_7.dct	r8_7.dta	Convert ascii file to stata file
	r8_8.do	r8_8.dct	r8_8.dta	Convert ascii file to stata file
	r8_9.do	r8_9.dct	r8_9.dta	Convert ascii file to stata file
	r9_1.do	r9_1.dct	r9_1.dta	Convert ascii file to stata file
	r9_2.do	r9_2.dct	r9_2.dta	Convert ascii file to stata file
INCOME				
	AgriculturalInputsAggregate.do	r6_6_2.dta	agriculturalinputsaggregate.dta	Aggregate costs for agricultural inputs
	CropAggregate.do	r6_6_1.dta strata.dta strata.dta	cropaggregate.dta	Aggregate income from crops
	DisabilityPensionAggregate.do	r1.dta r8_4.dta	disabilitypensionaggregate.dta	Aggregate disability pensions
	HouseholdBenefitsAggregate.do	r8_9.dta strata.dta strata.dta	householdbenefitsaggregate.dta	Aggregate household benefits
	IndividualInkindBenefitsAggregate.do	r1.dta r8_7.dta	individualinkindbenefitsaggregate.dta	Aggregate individula inkind (transport/medicine/etc) benefits
	LiveStockAggregate.do	r6_8.dta strata.dta deflator.dta deflator.dta	livestockaggregate.dta	Aggregate income from livestock
	MainJobAggregate.do	r1.dta r6_2.dta	mainjobaggregate.dta	Aggregate wage from main job
	OtherRevenuesAggregate.do	r8_10.dta strata.dta	otherrevenuesaggregate.dta	Aggregate revenues from miscellaneous sources
	PrivatePensionAggregate.do	r1.dta r8_2.dta	privatepensionaggregate.dta	Aggregate private pensions
	PublicPensionAggregate.do	r1.dta r8_1.dta	publicpensionaggregate.dta	Aggregate public (state old age) pensions
	RealEstateAggregate.do	r9_2.dta strata.dta strata.dta	realestateaggregate.dta	Aggregate income from real estates
	RemittancesReceivedAggregate.do	r7_1.dta strata.dta strata.dta	remittancesreceivedaggregate.dta	Aggregate remittances received from others
	RemittancesSentAggregate.do	r7_2.dta strata.dta strata.dta	remittancessentaggregate.dta	Aggregate remittances sent to others
	SecondJobAggregate.do	r1.dta r6_3.dta	secondjobaggregate	Aggregate wage from second job
	SelfEmploymentAggregate.do	r6_4_1.dta r1.dta	selfemploymentaggregate	Aggregate income from self-employment
	SelfEmploymentBusinessAggregate.do	r6_4_2	selfemploymentbusinessaggregate	Aggregate several variables for self-employment business
	SocialProgramAggregate.do	r1.dta r8_6.dta	socialprogramaggregate.dta	Aggregate maternity and child care system under social assistance system
	SurvivorsPensionAggregate.do	r1.dta r8_3.dta	survivorspensionaggregate.dta	Aggregate survivors pensions

	TotalIncomeAggregate.do	r1.dta strata.dta householdincometemporary.dta	individualincometemporary.dta individualincomeaggregate.dta householdincometemporary.dta householdincomeaggregate.dta	Aggregate hosuehold-level and individual--level income
	UnemploymentBenefitsAggregate.do	r1.dta r8_5.dta deflator.dta	unemploymentbenefitsaggregate.dta	Aggregate unemployment benefits
PROFILE				
	AssetProfile.do	r9_2.dta r6_8.dta r6_6_1.dta r4.dta	assetprofile.dta	Profile for each asset (own/not own)
	EmploymentProfile.do	r1.dta r6_1.dta	employmentprofile.dta	Create employment status(employed/unemployed/not economically active)
	HouseholdInformation.do	personalinformation.dta r4.dta householdincomeaggregate.dta povertyprofile.dta assetprofile.dta	householdinformation.dta	Collect various hosuehold-level information in one place
	PersonalInformation.do	r1.dta r3.dta employmentprofile.dta socialassistanceprofile.dta	personalinformation.dta	Collect various individual-level information in one place
	PovertyProfile.do	adjustedpercapitaconsumption.dta strata.dta	povertyprofile.dta	Create poor-nonpoor profile, depth and severity variables
	SocialAssistanceProfile.do	r1.dta r8_1.dta r8_2.dta r8_3.dta r8_4.dta r8_5.dta r8_6.dta r8_7.dta r8_8.dta	socialassistanceprofile.dta	Profile for each social assistance program (receive/not receive)