BASIC INFORMATION

PAKISTAN INTEGRATED HOUSEHOLD SURVEY
(PIHS) 1991

Poverty and Human Resources Division
The World Bank

December 1995

PRINCIPAL ABBREVIATIONS AND ACRONYMS USED

FBS	Federal Bureau of Statistics
HIES	Household Income and Expenditure Survey
LSMS	Living Standards Measurement Study
NWFP	Northwest Frontier Province
PIHS	Pakistan Integrated Household Survey
PPS	Probability Proportional to Estimated Size
PSU	Primary Sampling Unit

TABLE OF CONTENTS

1.0 Introduction:

The Pakistan Integrated Household Survey (PIHS) was conducted jointly by the Federal Bureau of Statistics (FBS), Government of Pakistan, and the World Bank. The survey was part of the Living Standards Measurement Study (LSMS) household surveys that have been conducted in a number of developing countries with the assistance of the World Bank. The purpose of these surveys is to provide policy makers and researchers with individual, household, and community level data needed to analyze the impact of policy initiatives on living standards of households.

The Pakistan Integrated Household Survey was carried out in 1991. This nationwide survey gathered individual and household level data using a multi-purpose household questionnaire. Topics covered included housing conditions, education, health, employment characteristics, self-employment activities, consumption, migration, fertility, credit and savings, and household energy consumption. Community level and price data were also collected during the course of the survey.

This document describes the design of the survey and its contents for potential users of the data. The sections that follow describe:

- Survey questionnaires
- Sample design for the survey
- Organization of the survey
- How to use the data
- Data quality

Additional information that is likely of interest to data users is contained in the appendices.

2.0 Survey Questionnaires:

The PIHS used three questionnaires: a household questionnaire, a community questionnaire, and a price questionnaire.

2.1 Household questionnaire:

The PIHS questionnaire comprised 17 sections, each of which covered a separate aspect of household activity. The various sections of the household questionnaire were as follows:

1. HOUSEHOLD INFORMATION
2. HOUSING
3. EDUCATION
4. HEALTH
5. WAGE EMPLOYMENT
6. FAMILY LABOR
7. ENERGY
8. MIGRATION
9. FARMING AND LIVESTOCK
10. NON-FARM ENTERPRISE ACTIVITIES
11. NON-FOOD EXPENDITURES AND INVENTORY OF DURABLE GOODS
12. FOOD EXPENSES AND HOME PRODUCTION
13. MARRIAGE AND MATERNITY HISTORY
14. ANTHROPOMETRICS
15. CREDIT AND SAVINGS
16. TRANSFERS AND REMITTANCES
17. OTHER INCOME

The household questionnaire was designed to be administered in two visits to each sample household. Apart from avoiding the problem of interviewing household members in one long stretch, scheduling two visits also allowed the teams to improve the quality of the data collected.

During the first visit to the household (Round 1), the enumerators covered sections 1 to 8, and fixed a date with the designated respondents of the household for the second visit. During the second visit (Round 2), which was normally held two weeks after the first visit, the enumerators covered the remaining portion of the questionnaire and resolved any omissions or inconsistencies that were detected during data entry of information from the first part of the survey.

Since many of the sections of the questionnaire pertained specifically to female members of the household, female interviewers were included in conducting the survey. The household questionnaire was split into two parts (Male and Female). Sections such as SECTION 3: EDUCATION, which solicited information on all individual members of the household (male as well as female) were included in both parts of the questionnaire. Other sections such as SECTION 2: HOUSING and SECTION 12: FOOD EXPENSES AND HOME PRODUCTION, which collected data at the aggregate household level, were included in either the male questionnaire or the female questionnaire, depending upon which member of the household was more likely to know more about that particular area of household activity. Male and female interviewers were instructed to switch questionnaires where necessary in order to obtain information from the best informed individual in the household.

Information for all male members aged 10 years or more was collected using the male questionnaire. Iinformation on other household members (i.e. all female household members as well as children aged less than 10 years) was collected using the female questionnaire. Individuals covered in the male questionnaire were assigned sequential ID codes beginning with code "01" and those household members covered in the female questionnaire were assigned ID codes starting with code "51".

It is important to note, however, that the division of the questionnaire into the male and female portions was undertaken solely to facilitate gathering of data in the field. Male and female enumerators could interview respondents of different sexes separately when visiting each household, and thus obtain information pertaining to household members of both sexes directly from the individuals concerned. This was particularly important in the case of sections such as SECTION 13: MARRIAGE AND MATERNITY HISTORY, where assigning female enumerators to

directly interview the women concerned was crucial. While information for male and female members was collected in separate questionnaires, these data were combined during data entry so that the household data files contain information on <u>all</u> members of the household.

Each section of the household questionnaire was further divided into subsections A, B, C, etc. A list of the subsections contained in each section, as well as the questionnaire part (i.e. male or female) in which it was placed, is as follows:

Table 1. The PIHS household questionnaires

SECTION	SUB-SECTION	PAGES	QUESTIONNAIRE

SECTION	SUB-SECTION	PAGES	QUESTIONNAIRE
12:	**FOOD EXPENSES AND HOME PRODUCTION:**	**4**	
	Part A: Food expenses	2	F
	Part B: Home production	2	F
13:	**MARRIAGE AND MATERNITY HISTORY:**	**6**	
	Part A: Maternity history for women 14 and older	1	F
	Part B: Family planning	1	F
	Part C: Maternity history for ever married women	2	F
	who have given birth		
	Part D: Infant feeding practices	1	F
	Part E: Men's marriage history	1	M
14:	**ANTHROPOMETRICS:**	**1**	F
15:	**CREDIT AND SAVINGS:**	**10**	
	Part A: Assets and liabilities position	1	M
	Part B: Borrowing and outstanding loans	2	M
	Part C: Lending and outstanding loans	1	M
	Part D: Property	1	M
	Part D1: Personal and investment property	1	M
	Part D2: Dowries	1	F
	Part D3: Stocks, shares, bonds, and other securities	1	M
	Part D4: Bank deposits and postal savings	1	M
	Part D5: Bisi or saving committees	1	M/F
16:	**TRANSFERS AND REMITTANCES:**	**2**	
	Part A: Remittances and transfer expenditure	1	M
	Part B: remittances and transfer income	1	M
17:	**OTHER INCOME:**	**1**	**M**

2.2 Community and price questionnaires:

In each of the 300 communities where household interviews were conducted for the PIHS, a community questionnaire was administered by the team supervisor. Respondents to this

questionnaire typically consisted of the head of the village or community, the local school master, local government official, or any other such individual who was knowledgeable about the community. Communities were defined as all households living in the Primary Sampling Unit (PSU) in which the interview was conducted (the concept of PSU is explained in more detail in the next section on Sample Design). While each of the 300 PSUs consisted of roughly the same number of households (generally about 200 - 300), the area covered by individual PSUs varied considerably. In urban areas, communities were, in general, much smaller in terms of area covered, and were defined to be the group of households living within the physical boundaries of the PSU. In rural areas, because of the low population density, the PSU at times consisted of a group of settlements spread over a large area. In such cases, the supervisors were instructed to treat the largest or most central village in the PSU as the community.

The community questionnaire contained questions on characteristics of the community such as the quality of physical infrastructure, provision of amenities such as electricity, gas and water, access to education and health care facilities, and on markets and availability of goods and services in the locality. In order to obtain more information on birth practices used in the community, one of the sections of the community questionnaire was directed at dais (birth attendants) in the community and contained a number of questions on birth practices and pre- and post-birth maternal care. In rural areas, in addition to the section on the general characteristics of the community, two additional sections on health facilities and primary school facilities were also administered. Detailed information was collected on the quality of infrastructure, the equipment and services available, as well as staffing of these facilities.

Finally, a price questionnaire was also administered in all the communities where households were interviewed. Price information for 37 goods was collected. The goods included items such as food staples, tea and sugar, selected vegetables, as well as a few non-food items like fuels, soaps, etc. For all goods, two sets of prices were collected: one from the local shopkeeper and the other from the local mandi or wholesale seller. In rural areas, prices of

agricultural inputs as well as other relevant information on local farming practices was also collected.

3.0 Sample design:

The sample for the PIHS was drawn using a multi-stage stratified sampling procedure from the Master Sample Frame developed by FBS based on the 1981 Population Census.

3.1 Sample frame:

This sample frame covers all four provinces (Punjab, Sindh, NWFP, and Balochistan) and both urban and rural areas. Excluded, however, are the Federally Administered Tribal Areas, military restricted areas, the districts of Kohistan, Chitral and Malakand and protected areas of NWFP. According to the FBS, the population of the excluded areas amounts to about 4 percent of the total population of Pakistan. Also excluded are households which depend entirely on charity for their living.

The sample frame consists of three main domains: (a) the self-representing cities; (b) other urban areas; and (c) rural areas. These domains are further split up into a number of smaller strata based on the system used by the Government to divide the country into administrative units. The four provinces of Pakistan mentioned above are divided into 20 divisions altogether; each of these divisions in turn is then further split into several districts. The system used to divide the sample frame into the three domains and the various strata is as follows:

(a) Self-representing cities: All cities with a population of 500,000 or more are classified as self-representing cities.   These include Karachi, Lahore, Gujranwala, Faisalabad, Rawalpindi, Multan, Hyderabad and Peshawar. In addition to these cities, Islamabad and Quetta are also included in this group as a result of being the national and provincial capitals respectively. Each self-representing city is considered as a separate stratum, and is further sub-stratified into low, medium, and high income groups on the basis of information collected at the time of demarcation or updating of the urban area sample frame.

(b) Other urban areas: All settlements with a population of 5,000 or more at the time of the 1981 Population Census are included in this group (excluding the self-representing cities mentioned above). Urban areas in each division of the four provinces are considered to be separate strata.

(c) Rural areas: Villages and communities with population less than 5,000 (at the time of the Census) are classified as rural areas. Settlements within each district of the country are considered to be separate strata with the exception of Balochistan province where, as a result of the relatively sparse population of the districts, each division instead is taken to be a stratum.

Table 2. Main strata of the Master Sample frame

| | PROVINCE | | | | |
DOMAIN	Punjab	Sindh	NWFP	Balochistan	PAKISTAN
Self-representing cities	6	2	1	1	10
Other urban areas	8	3	5	4	20
Rural areas	30	14	10	4	58
TOTAL	**44**	**19**	**16**	**9**	**88**

As the above table shows, the sample frame consists of 88 strata altogether. Households in each stratum of the sample frame are exclusively and exhaustively divided into PSUs. In urban areas, each city or town is divided into a number of enumeration blocks with well-defined boundaries and maps. Each enumeration block consists of about 200-250 households, and is taken to be a separate PSU. The list of enumeration blocks is updated every five years

or so, with the list used for the PIHS having been modified on the basis of the Census of Establishments conducted in 1988.

In rural areas, demarcation of PSUs has been done on the basis of the list of villages/ mouzas/dehs published by the Population Census Organization based on the 1981 Census. Each of these villages/mouzas/dehs is taken to be a separate PSU.

Altogether, the sample frame consists of approximately 18,000 urban and 43,000 rural PSUs.

Table 3. Primary sampling units (PSUs) selected for the PIHS

	PROVINCE				
D O M A I N	**Punjab**	**Sindh**	**NWFP**	**Balochistan**	**PAKISTAN**
Self-representing cities	**38**	**31**	**7**	**4**	**80**
- high income	10	5	2	-	17
- middle income	17	14	3	2	36
- low income	11	12	2	2	27
Other urban areas	**38**	**10**	**14**	**8**	**70**
Rural areas	**78**	**42**	**21**	**9**	**150**
T O T A L	**154**	**73**	**42**	**21**	**300**

3.2 Sample selection:

The PIHS sample comprised 4,800 households drawn from 300 PSUs throughout the country. The breakdown of PSUs by province and domain is presented in Table 3. Sample PSUs were divided equally between urban and rural areas, with at least two PSUs selected from each of the strata. Selection of PSUs from within each stratum was carried out using the probability proportional to estimated size method.[1] A list of the PSUs selected for the PIHS is presented in Appendix 1.

Once sample PSUs had been identified, a listing of all households residing in the PSU was made in all those PSUs where such a listing exercise had not been undertaken recently. Using systematic sampling with a random start, a short-list of 24 households was prepared for each PSU. Sixteen households from this list were selected to be interviewed from the PSU; every third household on the list was designated as a replacement household to be interviewed only if it was not possible to interview either of the two households immediately preceding it on the list.

As a result of replacing households that could not be interviewed because of non-responses, temporary absence, and other such reasons, the actual number of households interviewed during the survey - 4,794 - was very close to the planned sample size of 4,800 households. Moreover, following a pre-determined procedure for replacing households had the added advantage of minimizing any biases that may otherwise have arisen had field teams been allowed more discretion in choosing substitute households.

3.3 Sample design effects:

The three-stage stratified sampling procedure outlined above has several advantages from the point of view of survey organization and implementation. Using this procedure ensures that all regions or strata deemed important are represented in the sample drawn for the survey. Picking clusters of households or PSUs in the various strata rather than directly drawing households

---

[1]In urban areas, estimates of the size of PSUs were based on the household count as found during the 1988 Census of Establishments. In rural areas, these estimates were based on the populatino count during the 1981 Census.

randomly from throughout the country greatly reduces travel time and cost. Finally, selecting a fixed number of households in each PSU makes it easier to distribute the workload evenly amongst field teams. However, in using this procedure to select the sample for the survey, two important matters need to be given consideration: (a) sampling weights or raising factors have to be first calculated to get national estimates from the survey data; and (b) the standard errors for estimates obtained from the data need to be adjusted to take account for the use of this procedure.

3.3.1   Sampling weights:

If a simple random sampling procedure had been used to draw the sample for the survey, the data collected could have been used directly to obtain national as well as regional estimates without the need for sampling weights or raising factors. However, in using data from a sample drawn by the procedure outline above, allowance needs to be made for the fact that this sampling procedure does not give all households in the country an equal chance of being selected for the survey. If no sampling weights are used with the data, the resulting estimates are likely to be biased as different types of households may not be represented in the sample in the same proportion as they exist in the population as a whole

In simple terms, sample weights attempt to correct for the fact that different households in the country have different chances of being included in the sample for the survey. To allow adjustment to be made for over-sampling of certain strata in the PIHS sample, sampling weights have been calculated, and have been incorporated into the PIHS data sets that are distributed. These raising factors should be used to weight data in order to obtain nationally representative statistics. In what follows. The way these sampling weights have been calculated is briefly outlined below.

The first aspect of the sampling strategy adopted for the PIHS that needs to be taken into consideration when calculating sampling weights is the stratification of the sample frame. Instead of picking PSUs at random from the country as a whole, PSUs for the PIHS survey were selected so as to ensure that at least 2 were picked from each strata of the Master Sample frame. Half the sample was picked from strata in urban areas even though they constituted less than 32 percent

16

of the country's estimated population in 1991. In order to correct for such over-sampling, the weight for households drawn from each strata needs to include a component that is inversely proportional to the probability of selection of PSUs in that strata. In other words, the greater the assigned probability for selecting PSUs in a particular stratum, the lower the weight we should give to households picked from this stratum.

The second step of sample selection for the PIHS - i.e. the selection of PSUs within each stratum - was carried out using the probability proportional to estimated size (PPS) procedure. In this method, a large PSU is assigned a higher probability of selection than a smaller PSU by a factor that is directly proportional to their relative size. If an equal number of households are to be interviewed in each selected PSU, then this method in principle results in a self-weighted sample within each stratum. In other words, all households within the stratum have an equal chance of selection in the sample and should therefore be allotted the same weight. In practice, however, allowance almost always needs to be made for the fact that the actual size of the PSU as found during the household listing exercise differs from the estimated size on which the selection of the PSU from the sample frame was based. The weight assigned to households in different PSUs thus includes a second component that is directly proportional to the ratio of the PSU's actual size to its estimated size. Households in a PSU where the count during the listing exercise reveals the population to be 50 percent higher than that earlier supposed are thus given a weight 50 percent higher than that assigned to households in a PSU where these two counts are found to coincide.

Finally, the third step of sample selection - i.e. that of selecting households within each PSU - does not have any effect on sampling weights; therefore, all households within a particular PSU are assigned the same weight. This is because the "systematic sampling with a random start" procedure used to select households gives all households in the PSU an equal chance of selection. Even the use of replacement households in the case of the PIHS does not affect the assignment of weights within the PSU, as the process of selection of replacement households was the same as that used to select the other 16 households to be interviewed from the PSU.

The formula used to calculate the weight assigned to the various PSUs is as follows:

$$w_{ij} = k \times \frac{1}{p_{ij}} \times \frac{n_j}{s_j}$$

where w_{ij} is the weight assigned to households in PSU j of stratum i, k is some constant, p_{ij} is the assigned probability of selection of PSU j of stratum i, (i.e. the higher the given probability of selection, the lower the weight given to the PSU), n_j is the number of households in the PSU j as found during the listing exercise, and s_j is the number of households in the PSU j on which the PPS was based.

3.3.2   Calculation of Standard Errors:

The PIHS sample was designed to yield representative statistics at the national and urban/rural) levels. Care however should be taken when interpreting results for smaller analytic domains as the sample was not designed to be representative at a more disaggregated level. Thus, even with the use of the sampling weights, statistics for the smaller provinces such as Balochistan are likely to have high standard errors given the relatively small sample size in these domains. In this regard, it is important to note that when calculating standard errors for estimates derived from the PIHS data, allowance must be made for the fact that the survey used a multi-staged sampling procedure. Calculating standard errors using methods outlined in elementary statistical textbooks is likely to underestimate the true magnitude of errors as the techniques presented in these books often assume that simple random selection was used when drawing the sample.

In general, a multi-staged sampling scheme that involves picking a cluster of households at some stage is less efficient than one which involves simple random sampling. This is because neighboring households tend to have similar characteristics, and so a sample drawn from them reflects less of the population's diversity than a simple random sample of the same given size N. In such an instance, the standard errors associated with estimates based on data from a survey using a multi-stage stratified sampling procedure (such as the PIHS) will be higher than would be indicated by simple random sample-based statistical theory.

The magnitude by which the standard error would be underestimated if no allowance is made for the "cluster" effect depends on the characteristic being estimated. In general, the more homogeneous the households are within a cluster with respect to the characteristic being estimated, the less efficient a sampling scheme based on clustering, and the higher the true standard error of the estimate obtained. For most variables of interest, the degree of homogeneity within the cluster is likely to be low, and so the effect of ignoring the "cluster effect" when estimating standard errors is unlikely to be too serious. However, in some cases, the inter-cluster correlation with respect to the variable of interest may be quite large (for instance whether or not the household has electricity). In these cases, if no allowance is made for clustering, the magnitude by which the true standard error is underestimated will be high.

4.0 Organization of the survey:

4.1 Staffing:

Field work for the PIHS was carried out by 15 teams based at FBS regional offices throughout the country. Two teams each were stationed in Karachi and Lahore, while one team each operated out of the FBS offices in Peshawar, Bannu, Rawalpindi, Gujranwala, Faisalabad, Sargodha, Multan, Bahawalpur, Sukkur, Hyderabad, and Quetta.

Each field team consisted of 7 members; a supervisor (Statistical Officer), two male and two female interviewers (Statistical Assistants), a data entry operator (Key Punch and Verifying Officer), and a driver. The four interviewers were responsible for carrying out the household interviews under the supervision of the Statistical Officer in accordance with the timetable prepared for each team. While the rest of the teams traveled back and forth between the regional office and the PSUs where the interviews were conducted, the data entry operators remained at the regional offices throughout. In order to facilitate travel for the field teams, a vehicle was provided to each team for the duration of the survey.

Overall supervision and coordination of the field work was conducted by the PIHS management team based at the FBS office in Islamabad. During the initial phase of the project, technical assistance was provided to the PIHS management team by local consultants hired for the project. The PIHS management team consisted of six members: a Project Director, a Chief Statistical Officer, three Statistical Officers, and a Data Processing Manager. The team was headed by the Project Director who was responsible for administering the survey. He directed the work of the team and ensured the smooth running of the overall project. He was assisted in his duties by the Chief PIHS Section, and by the three Statistical Officers. The Data Processing Officer was responsible for working with consultants to develop the data entry software for the survey, and to ensure that the supervisors and data entry operators followed the instructions for running the programs and operating the microcomputers properly.

4.2 Schedule of activities:

Once preliminary arrangements regarding the outline of the project had been finalized, discussions were held between staff from the World Bank, the Federal Bureau of Statistics, Pakistani researchers, and donor agencies in order to develop a draft of the household questionnaire. This questionnaire was then field-tested in June 1990. Following the field test, a workshop was held in Islamabad where the FBS staff that had participated in the field work were invited to give their comments on the questionnaire. The household questionnaire was then revised and finalized in light of these discussions, and translated into Urdu.

Some of the field staff used for the PIHS were drawn from the personnel of the FBS, whereas the rest were recruited by the Bureau for the project. Training of the field staff was conducted in Islamabad during November and December 1990. Initially, a two week training session was organized for the team supervisors. The main topics covered during the course of this training were the organization of the survey and the supervisory checks to be performed on the work of the interviewers. The supervisors were then joined by the interviewers for the main training session. This session spanned four weeks; during the first three weeks, the field staff were given training on completing the household questionnaire itself while in the last week, the teams were taken to neighboring communities to conduct practice interviews. Supervisors were also able to practice supervisory checks during these visits. These household interviews were observed and critiqued by the survey staff.

Data entry operators received training for three weeks which was conducted concurrently with the training for the supervisors and interviewers. This training consisted of three main parts. First, as many of the trainees recruited for data entry had not used computers before, they were provided with training on the use and maintenance of personal computers. During the second part of the training, the data entry operators were instructed on the use of the data entry program. Finally, the training also included a practical training component where data entry operators recorded the data from the household interviews completed as part of the interviewer training. Printouts of the data entered were given to the team supervisors who then discussed the mistakes highlighted by the data entry program in these printouts with the interviewers concerned.

About 20 percent more staff than project requirements were trained during this period. This served two main purposes: (a) the project management team would use the most promising trainees for the main survey; and (b) the staff that dropped out during the survey or were unable to work temporarily could be replaced by the extra personnel that had been trained.

Following completion of the training in Islamabad, the various teams returned to their duty stations, and field work for the survey commenced in January 1991. During the course of the next twelve months, the PIHS field teams covered about 20 PSUs each on average. In the 300 PSUs covered, almost 4,800 households were interviewed.

4.3 Organization of field work:

The PIHS was the first survey conducted by FBS in which data entry was carried out directly in the field. The main reasons for conducting data entry in the field was to improve data quality (possible errors could be corrected in the field through revisiting the households concerning rather than carrying out office editing), and to reduce the time taken between the completion of field work and availability of data for analysis. Decentralizing the data entry process involved installing a microcomputer in each of the regional offices for the immediate entry of data from all questionnaires completed by each team.

The schedule of work for all teams consisted of completing two PSUs each in a four-week period. Each team completed the first round of interviews in PSU 1 during the first week, the first round of interviews in PSU 2 during the second week, returned to PSU 1 to complete the second round of interviews in the third week, and then completed the second round of interviews in PSU 2 during the fourth week. At the end of each week, the team returned to the regional office to give the questionnaires to the data entry operator for data entry. The schedule of household interviews and data entry is summarized in Table 4.

Table 4. Work Schedule of Field Teams

	WEEK 1	WEEK 2	WEEK 3	WEEK 4	WEEK 5
Field teams	PSU 1 Round 1	PSU 2 Round 1	PSU 1 Round 2	PSU 2 Round 2	
Data entry operator		PSU 1 Round 1	PSU 2 Round 1	PSU 1 Round 2	PSU 2 Round 2

As the table shows, data entry of interviews conducted in a particular week was carried out in the following week. Thus, before the team went back to any PSU for the second round, data entry of the first round for that PSU had been completed by the data entry operator. During the second round visit, teams could take with them printouts of the data entered from the first round with a record of data omissions, possible errors, and inconsistencies for correction or verification.

During a week, the team completed one round of interviews for 16 households in the PSU. The teams worked in two pairs of one male and one female interviewer each, with each pair covering on average 2 households per day. During the period when household interviews were being conducted, the team stayed in the PSU. On their return to the office at the end of the week, the supervisor would review the printouts of data from the households for possible interviewer and data entry errors. Data entry errors would then be corrected at the office, while other possible data errors or inconsistencies would be marked on to the questionnaires and given to the interviewers for correction during the next visit.

5.0 Using the data:

The data from the PIHS can be obtained on diskettes. In what follows, the first section briefly describes some of the documentation related to the PIHS that can be used to understand and decode the data. The next section explains how the PIHS data is organized. Section 3 describes in greater detail how observations in different data sets can be uniquely identified, as well as the information that is contained in each ID code. Finally, section 4 gives a brief introduction to how data users can merge together data from the various PIHS data sets to create data sets tailored to their needs. The procedure to follow to obtain the PIHS data is described in Appendix 2, while a list of supporting documents that might be of interest to data users is given in Appendix 3.

5.1 Data documentation:

The PIHS questionnaires - both the household as well as the community - used in conjunction with the dictionary of variables, provide the best sources for understanding and decoding the data. As will be explained in more detail in the next section, data for the PIHS is stored in many smaller files, each of which contains data from part or all of a page of the questionnaires. The questionnaires also contain the exact wording of the questions asked, as well as instructions to the interviewers, and so are very useful in interpreting the data.

Interviewers were directed to read out only the things written in lower case on the questionnaires, while upper case print was for instructions to the interviewers. Responses to all questions in the questionnaires were pre-coded and printed on the questionnaire (with the obvious exception of questions soliciting quantitative information). At times, the list of responses was to be read to the respondent, but more often the interviewer was simply to code the response given. All the codes corresponding to the questions asked of respondents are contained in the questionnaires themselves. The codes used for various questions are printed on the same page as the questions themselves, with the exception of a few such as the industry, occupation, and geographic codes which are listed at the end of the questionnaire.

In the survey questionnaires, extensive use was made of skip patterns in order to maximize the ease with which household interviews were conducted, and to minimize interview time. The structure of the skip patterns was designed to solicit all the desired information, but to allow the interviewer to exclude those questions that did not apply to that particular respondent or household. Data users must be aware of these skip patterns so that the data can be properly interpreted. In most cases, the skip pattern is very easy to follow. Unless otherwise indicated, the interviewer is to ask the respondent the next question. An arrow followed by a number in parenthesis (e.g. » 10) after a particular response indicates the next question which the interviewer should ask if that response is given. An arrow with a number in a rectangle indicates which question should be asked next, regardless of the response received.

Instructions to interviewers as well as definitions used for the purposes of the survey are described in detail in the interviewer manual. The complete list of codes for the PIHS data sets is also given in the PIHS dictionary of variables, which includes other information likely to be of use in understanding the data. For details on how to obtain these documents as well as copies of the survey questionnaires, please refer to Appendix 3.

5.2 PIHS data files:

The PIHS data is available on diskettes in SAS portable, Stata, or ASCII formats. The data are distributed in compressed form, and each set of diskettes contains the program necessary to decompress them. When decompressed, these data files have the file extension .SSP, .DTA, and .DAT respectively. The SAS and Stata files contain variable labels for most variables, while the ASCII data come with variable names only. In the description of the data that follows, reference is made in particular to the data distributed in ASCII format. However, since the Stata and SAS portable format data is organized in a similar way, much of the description is also likely to be of interest to analysts using these files.

5.2.1 PIHS household-level data:

The PIHS household data set is broken down into 184 data sets which are stored in separate files. Each of these data sets contains data from one page (or part of a page) of the questionnaire. The

name of the data file indicates which particular section of the questionnaire the data was obtained from. For instance, the file F01A contains data from SECTION 1: HOUSEHOLD INFORMATION: PART A: HOUSEHOLD ROSTER. Similarly, the files F04C1 and F04C2 contain data on questions 1-10 and 11-16 respectively of SECTION 4: HEALTH: PART C: OTHER ILLNESSES which covers two separate pages of the questionnaire. Data from the 17 different sections of the household questionnaire are thus divided into a total of 184 different files. These data files are assigned sequential record types ranging from 1 - 184, and these record types are the first variable included in each of the data files.

In many of the 184 data sets, identifying which household the data pertains to is a straightforward exercise - data for a household is stored in one observation (i.e. line) of these files. However, this is not the case with all 184 files. In some data files, data for each household is stored in a number of different observations. Moreover, in these files, the exact number of observations over which the data for a particular household is stored varies with each household. For instance, each household consists of different number of individuals. Thus, in the file F01A which contains data from SECTION 1: HOUSEHOLD INFORMATION: PART A: HOUSEHOLD ROSTER, information pertaining to a household which has 5 members is contained in 5 different lines, for one with 8 members in 8 lines, and so on.

Each observation of the data files contains information for a particular unit; as pointed out above, this could be one particular household, or one individual within the household, or any other such item of interest. In general, the level of observation for each data set - i.e. the unit to which the data pertains - depends on which section the data were obtained from. In the case of section 2, for instance, the unit of observation in the corresponding data sets F02A, F02B, and F02C is the household. In sections 3, 4, and 5 where questions are asked of many household members, the unit of observation for the corresponding data sets is each household member.

Similarly, in other data sets, various other levels of observation are used, such as each food code, each agricultural crop, etc. A list of the 184 PIHS data files, grouped together by level of observation, is given in Table 5. The first group in this table consists of those data files in

which each observation pertains to different <u>households</u> interviewed during the survey. The second group of files contain data organized at an <u>individual</u> level; in other words, data pertaining to each individual is stored as a separate observation in these files.

Table 5: Level of observation of the PIHS household data sets

Observation	ID code	PIHS data files							
1. Household	9 digits (HID)	00MA	00FA	02A	02B	02C	06A1	06B1	07FA1
		07FA2A	07FA3A	07FA4A	07FA5A	07FA5D	07FA6A	07FA6C	07FB1
		07FB5	07FC1	07FC5	07FD1	07FD5	07FE1	07FE2	07FE3
		07FF1	07FF2	07FF3	07FF4	07FG	07FH	07FI1	07FJ1
		07FK	07FL2	07FM1	07FM2	07MA1	07MA2	07MA3	07MA4
		07MB	07MC1	07MC2	07MD	07ME1	07ME2	07MF1	07MF2
		07MI2	07MM1	07MM2	09A1	09A2	09A3	09A4	09B4
		09C1	09D01	09D03	09D05	09D06	09D08	09D09	09D10
		09D12	09E1	09E2	09F1	09G1	09G3	09H1	09H3
		10A1	12B1	15A	15B1	15C1	15D0	15D1	15D5
		16A1	16B1						
2. Individual	11 digits (9 digit HID+ 2 digit PID)	01A	01B	03A	03B1	03B2	03C	03D	04A
		04B	04C1	04C2	05A1	05A2	05B1	05B2	05B3
		05B4	05B5	05C	05D1	05D2	06A2	06C	08
		13A	13B	13C1	13C3	13D	13E	14	
3. Line number	10-11 digits (9 digit HID+ 1-2 digit no:)	00MB	00FB	06B2	07FA2B	07FA2C	07FA2D	07FA3B	
		07FA3C							
		07FA3D	07FA4B	07FA4C	07FA4D	07FA5B	07FA5C	07FA6B	07FB2
		07FB3	07FB4	07FB6	07FC2	07FC3	07FC4	07FC6	07FD2
		07FD3	07FD4	07FJ2	09C2	09F2	09F3	09G2	11C
		15B2	15B3	15C2	15D2	15D3	15D4	16A2	16B2
		10A2	10A3	10A4	10C2	10D			
4. Crop	12 digits (9 digit HID+ 3 digit crop code)	09B1A	09B1B	09B1C	09B2A	09B2B	09B2C	09B3A	09B3B
		09D02	09D04	09D07					
5.Expenditure item	12 digits (9 digit HID+ 3 digit exp. code)	11A	11B	12A	12B2				

Observation	ID code	PIHS data files							
6. Other	10-12 digits	13C2	07FI2	07FL1	07MI1	06B3	09D11	09D13	09H2
	(9 digit HID+	10B	10C1	17					
	1-3 digt other								
	cd.)								

An important point to note with regard to the above table is the column labeled ID code. All households interviewed during the PIHS were assigned an ID code known as HID, a 9 digit code that is unique to each particular household. This ID code contains much useful information likely to be of interest to data users, and more information on its composition is given in Section 3. This HID is included in all observations, no matter which data file they are drawn from, and allows the data user to identify which household the observation pertains to.

Much useful information pertaining to the data in each of the 184 household data files is contained in the PIHS dictionary of variables. One page of this dictionary of variables is reproduced in Table 6 for reference. This page describes the data contained in F01A, the data file for SECTION 1: HOUSEHOLD INFORMATION: PART A: HOUSEHOLD ROSTER .

The first two variables listed on this page are HID and PID. The dividing line below these two variables indicates that these variables are sufficient to uniquely identify each observation. In the case of this data set, each observation corresponds to data on one household member. All remaining variables in the data set are also listed on this page, along with a description of the codes used for each question. The PIHS ASCII data distributed has now been reformatted, so that the information contained in the columns CODE, FROM, and LENGTH of the dictionary of variables no longer applies to the data layout. However, this document is very useful in that it helps users of the data connect variable names included with the data diskettes to questions in the survey questionnaire.

The first few lines of data from the ASCII file F01A are reproduced in Table 7. As the table shows, commas (,) are used as delimiters to separate data for different variables, and blanks in

the data are denoted by periods (.). The first line of the data file contains all the variable names. Each data file contains three sets of variables.

Table 6. Dictionary of Variables

RECORD 5: SECTION 1, PART A: HH INFORMATION - 1

```
┌─────────────────────────────────────────────────────────────────────┐
│      HH#: SECTION 1, PART A: HH INFORMATION      ID CODE:            │
│                                                                       │
│   2 SEX:                              _  °°°°°°°                      │
│                                                                       │
│   3 RELATIONSHIP WITH HEAD:           _  °°°°°°°°°°°°°°°°°°°°°°°       │
│                                                                       │
│   4 AGE IN YEARS:                     ___                            │
│                                                                       │
│   5 MARITAL STATUS:                   _  °°°°°°°°°°°°°°               │
│                                                                       │
│   6 SPOUSE LIVE AT HOME?:             _  °°°°                        │
│                                                                       │
│   7 ID CODE OF SPOUSE:                ___                            │
│                                                                       │
│   8 TIME AWAY IN MONTHS:              ___                            │
│                                                                       │
│   9 MEMBER OR NOT?:                   _  °°°                         │
└─────────────────────────────────────────────────────────────────────┘
```

VARIABLE	CODE	RT	FROM	LENGTH	TYPE	REMARKS
HID		5	4	9		
PID	IDC	5	13	2	QNT	VALUES RANGE FROM 1 TO 99
2 SEX	Q02	5	15	1	QLN	NOMENCLATURE: MALE....1 FEMALE..2
3 RELATIONSHIP WITH HEAD	Q03	5	16	2	QLN	NOMENCLATURE: HEAD...................01 WIFE OR HUSBAND........02 SON/DAUGHTER...........03 GRANDCHILD.............04 FATHER OR MOTHER.......05 SISTER OR BROTHER......06 NIECE OR NEPHEW........07 SON/DAUGHTER-IN-LAW....08 BROTHER/SISTER-IN-LAW..09 FATHER/MOTHER-IN-LAW...10 OTHER RELATIVE.........11 SERVANT/TENANT.........12 OTHER NOT RELATED......13
4 AGE IN YEARS	Q04	5	18	3	QNT	VALUES RANGE FROM 0 TO 120
5 MARITAL STATUS	Q05	5	21	1	QLN	NOMENCLATURE: MARRIED........1 DIVORCED.......2 (_Q08) SEPARATED......3 (_Q08) WIDOW/WIDOWER..4 (_Q08) NEVER MARRIED..5 (_Q08) (CAN BE BLANK)
6 SPOUSE LIVE AT HOME?	Q06	5	22	1	QLN	NOMENCLATURE: YES..1 NO...2 (_Q08) (CAN BE BLANK)
7 ID CODE OF SPOUSE	Q07	5	23	2	QNT	VALUES RANGE FROM 1 TO 99 (CAN BE BLANK)
8 TIME AWAY IN MONTHS	Q08	5	25	2	QNT	VALUES RANGE FROM 0 TO 12
9 MEMBER OR NOT?	Q09	5	27	1	QLN	NOMENCLATURE: YES..1 NO...2

Table 7. Data from F01A.DAT

```
CID,CLUST,NH,HID,PID,SEX,REL,AGEY,MAR,SCOHAB,SID,MOSABS,MEMB
5,1101001,1,110100101,1,1,1,37,1,1,51,0,1
5,1101001,1,110100101,2,1,3,17,5,.,.,0,1
5,1101001,1,110100101,3,1,3,13,5,.,.,0,1
5,1101001,1,110100101,51,2,2,41,1,1,1,0,1
5,1101001,1,110100101,52,2,3,14,5,.,.,0,1
5,1101001,1,110100101,53,2,3,8,.,.,.,0,1
5,1101001,2,110100102,1,1,1,43,1,1,51,0,1
5,1101001,2,110100102,2,1,3,20,5,.,.,0,1
......
```

The first set consists of three variables; CID, CLUST, and NH. CID refers to the record type of the data file (called RT in the dictionary of variables), CLUST to the PSU code, and NH to the number of the household. The next set, HID and PID (upper part of the dividing line in the dictionary of variables), are the variables that can be used to uniquely identify each observation. Finally, the third set of variables consists of data on the various questions that were asked in that particular section during the survey. Thus, the variable SEX contains data for question 2 of this section (sex of the household member), REL for question 3 (relationship to head of household), and so on.

The list of codes used is contained in both the dictionary of variables and in the household questionnaires themselves. A value of "1" for MAR thus denotes that the individual is "Married" (e.g. observation 1 in Table 7), while "5" indicates that the person's marital status is "Never married" (observation 2). In the dictionary of variables, codes 2-5 for the question on marital status are followed by (_Q08), thus indicating that following the skip pattern in the questionnaires, if this code is used in a particular observation, all variables up to Q08 "8 Time away in months" will be blank (e.g. observation 2). If a particular question was not to be asked of some household members, this is indicated by the statement (CAN BE BLANK) in the dictionary of variables. Thus, in observation 6 which pertains to data on the 8 year old daughter of the household head, variables MAR, SCHOHAB, SID all contain blanks. These questions were not to be asked of children less than 10 years age.

Data from other sections of the questionnaire can be decoded in a manner similar to that outlined above using the corresponding data files, the dictionary of variables, and the household questionnaires.

5.2.2   PIHS community-level data:

In addition to the household data files, data from the community interviews administered in each of the 300 PSUs visited by the PIHS teams are also available on diskette. These data are stored in 28 separate files named RT01.OUT - RT28.OUT.[2] As is the case with the household data sets, each of the data files contains data from one page, or part of a page, of the questionnaire.

The names of the files that contain data for each sub-section of the community questionnaire is given in Table 8. Data for questions 1 - 5 of the "Characteristics of urban communities" sub-section is contained in RT13.OUT, for questions 6 - 7 in RT14.OUT, and so on. The README.ASC file on the community data diskette describes in more detail the contents of each of the data files

Table 8: PIHS community-level data files

Community questionnaire section	Data files
1. Characteristics of urban communities	RT13.OUT - RT20.OUT
2. Characteristics of rural communities	RT02.OUT - RT12.OUT
3. Rural primary school questionnaire	RT25.OUT - RT27.OUT
4. Rural health facility questionnaire	RT21.OUT - RT23.OUT
5. Consumer price questionnaire	RT28.OUT
6. Questionnaire for Dais	RT24.OUT

The record layout of each of the data files is provided in the accompanying format files RT01.LST - RT28.LST. For example, RT13.LST, which is reproduced for reference in Table 9, describes the structure of the data contained in RT13.OUT. This data file contains 150 observations (corresponding to the 150 urban PSUs where the community interviews were conducted), each of which is 28 characters long. As the .LST files describes, the ID code of the

---

[2]As mentioned earlier, the data are available in SAS portable, Stata, and ASCII formats. The description here again pertains to the ASCII data files. Other formats, however, also follow a similar structure.

PSU starts at column 1, and is 7 characters long. Similarly, data for question 1: "Total persons residing in the PSU" occupies columns 8 - 13, data for question 2: "Number of households residing in PSU" occupies columns 14 - 18, and so on..

Table 9: Record layout of the community data

13.URBAN 1: POPULATION AND HOUSING CHARACTERISTICS				
RT	FROM	LENGTH	VARIABLE	NAME
13	1	7	1	PSU CODE
13	8	6	11	1. Total persons
13	14	5	12	2. Number of households
13	19	7	13	3. Average price Rs.
13	26	1	14	UNIT
13	27	1	15	4. Houses are ...
13	28	1	16	5. Streets are ...
RECORD LENGTH			28	
RECORDS PASSED			150	

5.3 Identifying observations:

As briefly mentioned earlier, each household interviewed in the survey was assigned a unique 9 digit identification code (HID) printed on the cover of each questionnaire. Observations in all PIHS data files contain this 9 digit household identification code, and this allows the data user to identify the household to which each particular observation pertains. Moreover, this 9 digit code also contains other useful information likely to be of interest to data users. The structure of this 9 digit code is as follows:

Table 10. Household identification code

1	2	3	4	5	6	7	8	9
PROVINCE	SUB-UNIVERSE	STRATUM		PRIMARY SAMPLING UNIT			HOUSEHOLD	

In the case of PROVINCE, possible values range from 1-4 (1: Punjab, 2: Sindh, 3: NWFP, 4: Balochistan). Thus all households which have an ID code beginning with 2 belong to Sindh. Similarly, in the case of SUB-UNIVERSE, possible values are 1 and 2, which indicate Urban and Rural areas respectively.

The first 7 digits of this ID code uniquely identify the cluster (i.e. PSU) from which the household was drawn. For instance, households with ID codes 113100201, 113100202, and 113100204, all belong to the same cluster in urban Punjab. In the case of the community data, the same convention is used to uniquely identify observations - i.e. for each PSU, the 7 digit PSU code used to identify observations in the community data files is the same as the first 7 digits of the HID assigned to households in that particular PSU. This greatly facilitates merging the household and community-level data. Merging together observations from different data sets is taken up in the next section.

Individuals within a particular household share the same household code. In addition to the 9 digit household code, each household member is assigned a 2 digit personal ID code (PID) ranging from 01 to 99 in Section 1A (Household Roster). By using the 11 digit ID obtained by combining the household ID code and the personal ID code, one can uniquely identify each person interviewed in the survey. For instance, by referring to this number, one can tell that individuals with ID code 11310020101, 11310020102, 11310020151, and 11310020152 all belong to the same household.[3] Similarly, observations in data sets F01A, F03A, F04C beginning with the same 11 digit ID code all pertain to data on the same individual.

In data files where there are multiple observations per household, all observations for a particular household share the same HID. Data in all sections of the questionnaire follows this same convention. For example, observations on different agricultural crops grown by a particular household have the 9 digit HID code, followed by a 3 digit crop code, observations on food items consumed by a particular household again share the same 9 digit household

---

[3] As mentioned earlier, PID 1-50 refer to male household members aged 10 years and above who were covered in the male questionnaire; PID 51-99 to female members and children who were covered in the female questionnaire.

code, followed by a 3 digit food item code, etc. Thus, by referring to this code, data users can identify the household to which the data pertain.

5.4 Merging data from different data sets:

The convention of assigning unique household, personal, and other ID codes helps greatly in allowing data users to merge together data from different parts of the questionnaire, and to create data sets tailored to their particular interests. For instance, an analyst interesting in studying the data on education could create a primary data set combining data from files F03A, F03B1, and F03B2. Data from these files could be merged together using the HID and PID codes.

A child in the household can be linked to the parents, if they are members of the household, through the ID codes of parents in SECTION 1B. For parents that are not members of the household, this section contains information on their level of education as well as the main occupation in which they were primarily engaged. Information on the spouse of a particular member of the household can be linked by first finding out the ID code of the spouse from SECTION 1A, and then using this ID code to obtain the necessary information for the person concerned from the various sections. Similarly, data at the aggregate household level (for instance, total household income) can be merged with individual-level information using HID as the merge variable.

Data on the household weights is contained in the data set WEIGHTS.DAT. This data set consists of two variables, CLUST and WEIGHT, which refer to the PSU visited during the PIHS and the associated weight (i.e. raising factor) respectively. This data set can be merged with other household data sets using the CLUST variable contained in them.

Data from the community questionnaire can also be merged with household or individual-level data sets. The analyst should first create a PSU CODE variable using the community data set. This variable can then be used to merge information from the community-level data files by matching it with the CLUST variable in the household data files.

Using the procedure outlined in the various examples given above, data users can combine data from various parts of the questionnaire to create a primary data set suitable for their particular needs.

5.5 PIHS data constructed aggregates:

A list of some of the research papers prepared using data from the 1991 PIHS is provided in Appendix 4. In preparing some of these papers, individual researchers have constructed household income and expenditure aggregates for their own purposes. These researchers have made these aggregates available, as well as the programs used to construct them, to other data users.

These data sets have been placed in the public domain to allow others who wish to use them to save on the time and effort required to recreate these aggregates. Potential users should note that the authors concerned have provided this information on the explicit understanding that: (a) they disclaim any responsibility for errors or mistakes that may unintentionally have been made in constructing these aggregates; and () no further information or explanation of the process by which these aggregates were constructed will be provided, other than that already contained in the accompanying documentation.

Two such aggregate data sets are available. The data set PIHSXPN.xxx contains data on aggregate household expenditure, and was constructed in two stages. In the first stage, the various components of aggregate expenditure were identified and computed, and were brought together in one data set. In the second stage, the variables were re-coded to give a coding structure compatible with the FBS Household Income and Expenditure Surveys (HIES). See Appendix 6 for more information.

Another data set, PIHSINC.xxx, contains data on aggregate household income. This data set was also constructed in two stages. In the first stage, income from various sources such as wage employment, agriculture, family enterprise activities, etc. was computed and stored in several data sets. In the second stage, income from these various data sets was brought together into one

aggregate household level data set. See Appendix 5 for more information.

Those wishing to use these data sets should be aware that computation of these expenditure and income aggregates has, in many instances, involved making explicit choices between a number of possible options. For instance, in the case of aggregate expenditure, food prices are required to convert quantities received in kind to values. There are a number of sources that could be used to obtain these prices, and the actual sources used in this case reflect the preference of the authors concerned. Similarly, in the case of aggregate income, a measure of the imputed rental value of owner-occupied housing was included in household income, which may or may not be in accordance with the definition of income that a particular analyst would like to use. To some extent, the accompanying documentation points out the places where such methodological choices have been made, and so users could, if they wished, amend the programs to reconstruct these data sets in line with their particular preferences.

6.0 Data quality:

Evaluating data quality is essentially a subjective matter and is therefore perhaps best left to users to assess for themselves. Unlike data from other surveys in Pakistan, the PIHS data that are distributed are essentially in the same form as were received from the field. In other words, no "office editing" or "cleaning" of the data has been carried out. Data users can judge for themselves the quality of the data, and are free to choose their own particular method for cleaning or correcting the data.

In the PIHS, information was collected on a wide range of topics, and the quality of data collected varies considerably between the different sections. For instance, a number of problems were found with the data from the anthropometrics section where analysis of the data suggested that measurement and recording errors had reduced data reliability considerably.[4] On the other hand, the quality of data collected on employment and economic activities is good. Even though women's labor force participation rates obtained from the PIHS are considerably higher than those obtained from earlier surveys, there are strong grounds for trusting the PIHS estimates. A more comprehensive approach to collecting information on employment was used in the survey than in earlier surveys. In addition to the sections on wage employment in agriculture and non-agriculture, the survey questionnaire also included detailed sections on on-farm and off-farm economic activities. Moreover, the use of female interviewers in the field who could interview women directly (unlike previous surveys) meant that information on women's employment was obtained first-hand.

A decentralized system of data entry was used in the PIHS which helped greatly in improving data quality. Use of special data entry software resulted in early detection of possible mistakes in the field where they could be rechecked by returning to the concerned households and corrected where necessary. In this section, the types of checks used in this software to improve data quality are described briefly in the first part. Some of the types of systematic problems that

---

[4] Kees Kostermans (1994): Assessing the Quality of Anthropometric Data LSMS Working Paper No: 101, Washington, The World Bank.

have been found in the PIHS data are then described in the second part. Finally, a comparison of the PIHS with other surveys is briefly summarized in the third part.

6.1 PIHS data entry program:

The methodology of decentralized data entry in the field used for the first time in Pakistan in the PIHS helped greatly in improving data quality. Early detection of possible errors by the data entry software allowed survey teams to check possible mistakes and, where necessary, to correct them when they revisited the households.

The data entry program used for the survey was designed to check for data entry errors, coding mistakes, as well as to search for incomplete or careless data collection by the interviewers. All the possible codes used in the questionnaire were incorporated into this software, and this helped reduce coding and data entry errors. If, for instance, the data entry operator entered the code "3" by mistake for a question for which the only admissible response was "Yes" or "No" (i.e. code 1 or 2), the program would alert the operator to the mistake by making a loud beep. The operator could then check his work more closely and correct any mistake made during data entry. In addition to alerting the data entry operator, the program also highlighted such coding mistakes in the printout of the data entered for the household. Thus any such errors missed by the data entry operators could be spotted by the supervisor and corrected.

In all places where quantitative information was sought in the questionnaire (amount received as income, quantities purchased, expenditures, etc.), the program also contained range checks. If the data entry operator entered data which was outside the bounds of the programmed range checks, the program would alert him. In such cases, operators were advised to check to see if the data entered in the computer matched the information filled into the questionnaire by the interviewers and, where necessary, to correct data entry mistakes. However, if the data entered into the computer matched the information provided in the questionnaires, the data entry operators were instructed not to make any changes to the data entered. Such out of range values that remained were automatically highlighted by the program in the printout of the data, and thus brought to the supervisor's attention. If necessary, the supervisor would then instruct the interviewers to recheck the information during the revisit to the concerned household. For

example, the upper range for the value of the respondent's house (Section 2, question 12) was set to Rs. 2 million. This particular value was set, not because this was considered to be the highest possible value reported by the households, but rather to minimize the number of households for which value of housing was overestimated. When such possible errors were highlighted by the program, the supervisor would then decide if the value was reasonable for the household concerned or, if necessary, instruct interviewers to recheck this bit of information during the revisit to the household.

Finally, the data entry program also contained a series of checks to ensure that the data collected for a particular household were internally consistent. The skip program used in the questionnaire was programmed into the data entry software. The program would check to ensure that the data entered conformed to the desired skip pattern, and that the interviewers or data entry operator had completed all the necessary questions. For instance, if the household reported having purchased a particular good, the program would check to see if necessary information on quantities purchased and expenditure were also recorded. Similarly, the program would also check to see if data in the various sections were collected for all eligible members of the household, and would alert the supervisor to any members that had accidentally been missed during the first round of data collection.

These pre-programmed checks in the data entry software coupled with decentralized data entry greatly improved the quality of data collected. This, among other reasons, was why data collected in the PIHS wer available for analysis very soon after the survey was completed without having to wait for extensive office editing of the data.

6.2 Data problems:

As mentioned above, the PIHS data that are distributed have not undergone any pre-release editing or cleaning, and data users should be aware of some of the types of problems that remain in the data they receive.

Despite the checking of possible coding errors in the data by the data entry software, a number of mistakes remain in the data. In some cases, resolving the problems that arise as a result is relatively easy. For example, at times the interviewers did not follow the skip pattern outlined in the survey questionnaires, and asked questions of the respondents that did not apply to those particular households or individuals. In such instances, by reference to the skips printed in the questionnaire and highlighted in the dictionary of variables, the data user can identify the "incorrect" data and amend it accordingly.

In other cases, however, correcting the coding mistakes is not so straight-forward. For instance, by incorrectly using the code "2" instead of "1" for unit of measurement, the interviewer (or data entry operator) may have misreported a household's weekly purchase of vegetables to be 2 maunds instead of 2 kilos. Or a person's wage earnings may have incorrectly been reported as 2000 rupees a day instead of 2000 rupees per month. Such mistakes are usually easy to spot as they show up as outliers in the data. However, correcting them is much more problematic as it essentially involves making some strong assumptions regarding the nature of mistakes made. In the example given above, one could decide that 2 kilos/week was a much more reasonable amount than 2 maunds/week and was therefore the correct response. However, this involves making a personal judgment; different users' notion of what is "reasonable" may differ considerably. In order to avoid making any assumption on the behalf of others, such types of possible errors have been left unedited in the data distributed. Analysts should be aware that these problems remain in the data sets they receive so that they can clean data according to their own criteria.

Another type of data problem pertains to missing values -- information that should have been asked for a particular household, but has been left out for one reason or the other; in contrast to those questions which, because they were not relevant, interviewers were instructed by the skip pattern to omit. For instance, for food codes 333-335 in the Urdu questionnaire, the question on value of purchased consumption was inadvertently blacked out. Interviewers were instructed nevertheless to ask the question. Some did, others didn't. As a result, data for this particular variable contain an exceptionally large number of missing values. Similarly, the question on imputed value of owner-occupied housing in the HOUSING section has a fairly high number of

missing values - no doubt because a number of respondents did not feel able to answer this question.

In attempting to remedy problems caused by such missing data, one could run a regression to estimate the household's consumption of these goods, given other expenditures, or the "expected" value of rent, given the various housing characteristics. However, as using such procedures involves an element of arbitrariness and personal preferences, missing values in the data have been left unchanged in the data sets distributed.

By necessity, for a large and complex survey such as the PIHS, any discussion or list of problems with the data can only be partial in coverage. By pointing out some of the main types of problems that users may encounter, as well as highlighting the fact that the released PIHS data have not been subject to office editing or cleaning, this section seeks mainly to make users aware of some of the data problems that they may encounter during analysis. Wherever required, each user can then make adjustments and clean the data according to his or her own particular methodology.

6.3 Comparison with other surveys:

A more detailed comparison of the PIHS with the 1984-85, 1987-88, and the 1990-91 HIES has been undertaken elsewhere.[5] To summarize briefly the main findings, the PIHS was found to contain households of similar age and gender structure as the other surveys, but of larger size on average. In general, household heads in the PIHS were found to be better educated than those in the 1990-91 HIES, especially in rural areas. Average household consumption in the PIHS was also found to be considerably higher than the other surveys. However, differences in estimates obtained from the various surveys are partly to be expected as these surveys use different definitions and methodologies.

---

[5] Howes, S and Zaidi, S (1994): Notes on some household surveys from Pakistan in the eighties and nineties. mimeo, STICERD, London School of Economics.

On the issue of representativeness - i.e. which of the surveys gave a more accurate picture of the country - it is difficult to make a definitive statement. Neither the HIES nor the PIHS come close to replicating the household size estimate of the Census. On the consumption side, the only outside source to compare estimates from the surveys with is the National Accounts. Estimates of household consumption from the HIES surveys are more similar to one another, thus suggesting that the PIHS may have overestimated consumption. However, estimates of household consumption from the PIHS are much closer to those derived from National Accounts statistics. Further, data for the PIHS reveal a much higher dispersion of income compared to the HIES surveys, especially for high income groups whose income the HIES surveys have frequently been criticized for underestimating.

APPENDIX 1: LIST OF PIHS PRIMARY SAMPLING UNITS:

1. PUNJAB:

a) Self-representing cities:

1131002	Lahore	1131014	Lahore
1131021	Lahore	1132002	Lahore
1132013	Lahore	1132073	Lahore
1132078	Lahore	1132118	Lahore
1132124	Lahore	1132182	Lahore
1132183	Lahore	1133027	Lahore
1133033	Lahore	1141001	Faisalabad
1141024	Faisalabad	1142016	Faisalabad
1142038	Faisalabad	1142047	Faisalabad
1143001	Faisalabad	1143010	Rawalpindi
1151003	Rawalpindi	1151010	Rawalpindi
1152025	Rawalpindi	1152045	Rawalpindi
1153011	Rawalpindi	1153019	Rawalpindi
1161004	Multan	1161010	Multan
1162007	Multan	1162015	Multan
1163011	Multan	1163020	Multan
1171001	Gujranwala	1171008	Gujranwala
1172005	Gujranwala	1172015	Gujranwala
1173011	Gujranwala	1173018	Gujranwala

b) Other urban areas:

1101001	Attock MC	1101016	Islamabad MC
1101028	Wah Cantt.	1101061	Dina TC
1101089	Chakwal MC	1102001	Mandi Bahauddin MC
1102027	Kotli Loharan MC	1102068	Pasrur MC
1102101	Kamoke MC	1103004	Kot Moman TC
1103029	Bhakkar MC	1103034	Phularwan TC
1103068	Sargodha MC	1103108	Mianwali MC
1104005	Thandlianwala TC	1104041	Jhang MC
1104062	Chiniot MC	1104105	Gojra MC
1105001	Kahna Nau TC	1105018	Kot Radhakishan TC
1105034	Sangla Hill MC	1105041	Kasur MC
1105063	Shahkot TC	1105080	Basirpur TC
1105108	Haveli Lakha TC	1106006	Sahiwal MC
1106014	Vehari MC	1106073	Shaujabad MC
1106094	Abdul Hakim TC	1107001	Dera Ghazi Khan MC
1107029	Jampur TC	1107043	Karor TC
1107050	Alipur TC	1108005	Bahawalpur MC
1108038	Ahmedpur East MC	1108054	Rahimyar Khan MC
1108111	Donga Bonga TC	1108114	Khanpur MC

c) Rural areas:

1201001	Panjgran	1201004	Kot Hathial
1202040	Basal	1202052	Dhak
1203003	Turkwal	1203034	Nambal
1204002	Sultan Pur	1204016	Pinanwal
1205042	Tarap North and South	1205053	Pira Fattiall
1206017	Dharema	1206024	Bhadhra
1207017	Okhli Mohla Janubi	1207030	Adhikot
1208017	Wan Bhachran Janubi	1208019	Sultan Khel Gharbi
1209019	Dhingana	1209024	Dhandala
1210067	Chak 441/GB Sadhora	1210087	Chak 097/RB Johal
1210103	Chak 451/GB Sado Anna	1210115	Chak 264/RB Nag Khurd
1211001	Chak 303/JB Katohar Kalan	1211026	Chak 189/GB Ardor Abad
1212008	Rodu Sultan	1212043	Kaki Nau Doim
1212054	Warh Thatta Mohd. Shah	1212087	Doka Baluchan
1213007	Garmula	1213034	Wanian Wala
1213078	Mardexe	1213080	Kaulo Tarar
1214002	Pandowal Bala	1214033	Chorund
1214067	Chak Sada	1214091	Khohar
1215001	Banbajwah	1215020	Rajian
1215042	Adalat Garh	1215107	Fatowal
1216008	Manga Utar	1216016	Hanjar Wal
1217007	Tal Wandi	1217036	Dhing Shah
1217041	Bharwal Kalan	1218030	Sharaq Pur Khurd
1218063	Rahan Wala	1218108	Chak No 175/RB
1219014	Quila Dev Singh	1219035	Kohla
1219049	042/SP-Samundri	1220004	Chak No 098/EB
1220034	Chak No 228/EB	1220043	Karam Pur
1221005	Chak No 169/9L	1221064	Chak No 059/EB
1221106	Muhammadpur	1222019	Qasba Sani
1222062	Gogran	1222082	Sikandarabad Gharobi
1223045	Chak No 132/10 R	1223067	Chak No 127/15 L
1224006	D.J.K. Darmiani	1224016	Nutkani
1225010	Chak Tariqabad	1225035	Sikhani Wala
1226042	Qalandar Wala	1226059	Ghulam Ali Gharbi
1227003	Warasehran	1227013	Nawan Kot Gharbi
1228008	Maushera Jadid	1228035	Dera Masti
1229002	Sayd Sharkanwali	1229032	Chak No 213/Fateh
1229067	Hasan Wala	1230021	Kot Karam Khan
1230044	Goth Mahi	1230055	Sanjarpur Nao

2. SINDH:

a) Self-representing cities:

2181001	Karachi	2181004	Karachi
2181021	Karachi	2181063	Karachi
2181084	Karachi	2181122	Karachi
2181127	Karachi	2181134	Karachi
2181193	Karachi	2181196	Karachi
2182014	Karachi	2182018	Karachi
2182041	Karachi	2182072	Karachi
2182077	Karachi	2182087	Karachi
2182137	Karachi	2182144	Karachi
2182151	Karachi	2182204	Karachi
2182219	Karachi	2182238	Karachi
2183005	Karachi	2183038	Karachi
2183060	Karachi	2191005	Hyderabad
2191006	Hyderabad	2192001	Hyderabad
2192015	Hyderabad	2193015	Hyderabad
2193017	Hyderabad		

b) Other urban areas:

2101029	Setharja TC	2101031	Ghotki TC
2101048	Nawabshah MC	2102001	Jacobabad MC
2102036	Kambar MC	2103004	Tando Allahyar MC
2103005	Dadu MC	2103044	Tando Adam MC
2103050	Badin MC	2103061	Mirpur Khas MC

c) Rural areas:

2201008	Shah Ladhani	2201013	Setharjaupper
2202014	Sangi Ghotki	2202030	Dad Loi
2202046	Begmanji	2203015	108/Nusrati
2203038	Khinyardon	2203043	Bao
2203054	Panhwar	2204005	Baragh
2204020	Dasti	2204038	Misri Pur
2205003	Kandhar	2205029	Lali Old
2206008	Daragad	2206022	Lakha
2206038	Faridabad	2207003	Baghban
2207027	Railo	2207042	Radhan
2208001	Hala New	2208009	Metkhan
2208029	Lankhiar	2208044	Khutiro
2208075	Singhr	2209019	Dei Jarkas

2209032	Kand Rakhi	2209040	Kario I&II
2210005	Samathri	2210016	Hingorno
2210033	Lundo	2211005	Todri
2211017	Shakhro	2211025	Akuto
2211036	Kinjheji	2211055	Deh 305
2211068	Melan Har	2212010	Kohistan 7/1
2212016	Duhro	2212029	Jhoke
2213015	Rehri	2214009	Manghopir

3. NWFP:

a) Self-representing cities:

3121001	Peshawar	3121010	Peshawar
3122008	Peshawar	3122012	Peshawar
3122019	Peshawar	3123011	Peshawar
3123020	Peshawar		

b) Other urban areas:

3101001	Mingora	3101017	Mingora
3102001	Mardan MC	3102011	Charsada MC
3102022	Sawabi MC	3102028	Nowshera MC
3102036	Topi TC	3103001	Kohat MC
3103030	Karak TC	3104017	Bannu MC
3104022	D.I. Khan MC	3105015	Baffa TC
3105034	Abbotabad MC	3105053	Havelian TC

c) Rural areas:

3201001	Kokarai	3201032	Chagam
3201060	Anghapur	3202002	Ali Gasar
3202061	Sadbar Kalai	3203002	Shergarh
3203041	Mathni Chungun	3204067	Razar
3204091	Landi Akhun Ahmed	3205001	Mohd. Khowja
3205011	Mohd. Zai	3206014	Bahadar Khel
3206023	Thati Nasrati	3207001	Fatima Khel Kalan
3207017	Dadiwala	3208017	Daraban
3208024	Lundah	3209002	Kaghan
3209048	Behali	3210050	Dewal Manal
3210082	Sobra		

4. BALOCHISTAN:

a) Self-representing cities:

4111006	Quetta	4111010	Quetta
4112024	Quetta	4112039	Quetta

b) Other urban areas:

4101001	Pishin MC	4101044	Zhob MC
4102003	Sibi MC	4102015	Dera Murad Khan Jamali
4103016	Khuzdar MC	4103027	Bela TC
4104019	Ormara TC	4104032	Turbat MC

c) Rural areas:

4201013	Station Musakhel	4201054	Vila Akarin
4201074	Ali Zai	4202039	Taib
4202045	Manjhooti	4203003	Moli
4203015	Hassanzai	4204009	Lebnan
4204033	Sarwan		

APPENDIX 2: OBTAINING THE 1991 PIHS DATA:

The 1991 PIHS data are the property of the Pakistani Government. In 1994, the Federal Bureau of Statistics adopted a policy of making the data freely available to researchers. Those who want to obtain the data should write to:

> The Chief
> PIHS Section
> Federal Bureau of Statistics
> G - 8 Markaz, Islamabad
> Pakistan

Alternately, the data can be obtained from:

> Living Standards Measurement Study
> Poverty and Human Resources Division
> Policy Research Department
> The World Bank
> 1818 H Street, N.W.
> Washington D.C. 20433
> USA

For those seeking to obtain the data through the World Bank, the letter should include a 1-2 page description of the proposed research to be undertaken using the data. There is a nominal fee associated with the data, which are available on diskette, in SAS portable (version 6.08), Stata (version 2.1), or ASCII files.

Copies of all reports and documents resulting from research on the data must be provided to the Federal Bureau of Statistics of Pakistan and the Poverty and Human Resources Division of the World Bank.

The researcher should further note that once received, the data cannot be passed on to a third party for any reason. Other researchers must contact the Federal Bureau of Statistics of Pakistan or the World Bank directly for access to the data. Any infringement on this policy will result in the denial of future access to World Bank data.

APPENDIX 3: LIST OF SUPPORTING DOCUMENTS:

The following documents can be obtained from the World Bank Poverty and Human Resources Division, at a cost of 0.05 cents per page for photocopying. All documents are available in English. The Household Questionnaires are also available in Urdu.

1.   1991 PIHS Male and Female Questionnaires (91 and 82 pages respectively)

2.   1991 PIHS Community Questionnaire (20 pages)

3.   1991 PIHS Interviewer Manuals
              Part I: Field Operations (45 pages)
              Part II: Household Questionnaires (123 pages)

4.   1991 PIHS Supervisor Manuals
              Field Operations (54 pages)
              Community-level questionnaires (16 pages)

5.   1991 PIHS Data Entry Manual
              Instructions for KPVOs (39 pages)

6.   1991 PIHS: Dictionary of Variables (315 pages)

7.   1991 PIHS Final Results (100 pages)

APPENDIX 4: LIST OF REPORTS/PAPERS USING 1991 PIHS DATA:

Gazdar H., Howes S., and Zaidi, S. (1994)
Recent Trends in Poverty in Pakistan. Mimeo, STICERD, London School of Economics.

Kees, Kostermans (1994)
Assessing the quality of anthropometric data LSMS working paper No: 101. World Bank, Washington.

Howes, S. and Zaidi, S. (1994)
Notes on some household surveys from Pakistan in the eighties and nineties. Mimeo, STICERD, London School of Economics.

Pritchett, L. and Filmer, D. (1996)
Environmental Degradation and the Demand for Children: Searching for the Vicious Circle. Draft. Poverty and Human Resouces Division, Policy Research Department, World Bank.

Schaffner, J.A. (1995)
Labor Markets in Developing Countries: Policy-Relevant Research Agendas and Implications for Household Survey Designs. Stanford University, Stanford, California.

Zaidi, S. (1993)
Demand for Housing and Urban Amenities in Pakistan. M. Phil. Thesis, Nuffield College, Oxford University.

APPENDIX 5: NOTES ON THE PIHS INCOME AGGREGATES:

Included with the raw data is a file containing an aggregated income variable and all of the program files (as well as the intermediary data files) used to construct it. The consultant who constructed this aggregate has agreed to allow the distribution of this information with the understanding that the descriptions given in this document and in the program files are the only documentation that will be provided. The aggregated income file can be found in a compressed file called INCOMxxx.ZIP, where xxx is DAT, DTA, or SSP, depending on whether the user has requested to receive ASCII, Stata, or SAS files. Decompressing this file results in three files--the income data file (PIHSINC.xxx), the income program file (income.sas), and another compressed file (BACKGRND.ZIP) which archives all of the program and data files. All of the materials in BACKGRND.ZIP have been prepared using SAS/PC. None of these files have been converted for use in Stata or as ASCII format. The contents of these three files are described in more detail below.

Any manipulation of the data requires that assumptions be made and, to the extent possible, those assumptions are explained below. Given the complexity and detail involved in the different income modules, it is possible to construct an income aggregate in different ways. Any researcher not satisfied with the assumptions made in this income aggregate should build their own estimate from the household data or alter the program files provided with this income aggregate. The following briefly outlines the process by which the estimates of household income and its components were calculated from the PIHS data sets.

Income aggregates for the PIHS were prepared using 22 separate programs. (See table below for the list of the programs. All of these program files can be found in the compressed file BACKGRND.ZIP.) Each of the programs starts with a primary SAS data files (F01A, F01B, etc.), and computes income earned by the household (or individual as the case may be) from that particular source. For instance, the file W_AGRI.SAS computes income earned from wage employment in agriculture by household members. It produces a SAS data set called W_AGRI in which income earned by family members during the past 12 months is stored in a variable called W_AGRI. (This file can be found in the data subdirectory of the BACKGRND.ZIP file.)

In this case, each observation in the data set corresponds to <u>individuals</u> who earned income from this particular activity. In other cases, each observation corresponds to the income earned by each <u>household</u> from that particular activity.

Observations pertaining to households can be identified by a unique household code named HID in the data sets. In the case of data sets where individuals are the unit of observation, each observation can be uniquely identified by the combination of the household code HID and the individual code PID.

Details concerning the various steps taken in computing income from various sources, as well as the assumptions made, are all documented in the respective SAS files. If data users would like to use a different definition of income, or proceed using different assumptions, they can modify the programs concerned accordingly.

In a few cases, information from other sections was drawn where necessary to calculate income (e.g. in-kind payments in wage agriculture) or where data from the section were deemed unreliable (e.g. income from sugarcane where prices from the community questionnaire were used in preference to the prices reported by households). In some cases, income had to be estimated as the data did not permit explicit calculation of income for some of the households (e.g. housing section where rents were imputed for some households). Finally, in some cases (e.g. income from family enterprise activities), outliers were replaced with values deemed more appropriate. At times, the assumptions used are based on little more than an educated guess, and can be criticized in some cases as being quite arbitrary. However, in all the above cases, the steps used and assumptions made are documented in the SAS programs so that users can modify these programs to suit their needs or preferences.

If the researcher wants to re-generate a measure of household income, he/she may choose to either start from 'scratch' or alter the included SAS programs and re-run them. In order to re-run the programs, all the primary data files for the PIHS (i.e. the 184 files prefixed F....) should be converted to SAS/PC data sets and stored in a directory with LIBNAME P91. (The programs were written for SAS/PC version 6.08. Any change from this platform or version may require

small changes to the programs.)   Users should assign the LIBNAME X. to the directory where they would like the generated data sets kept. In order to use these programs, two additional files should be included in the directory with the primary SAS files: one is called TRACTOR.SD2 which contains data on tractor rental rates from the community questionnaire (in a variable called T_RENT), and the other is an ASCII file of the data. (The program AGRINPUT uses this ASCII file.)

If the researcher is satisfied with the assumptions made to construct the measure of household income, all he/she need do is decompress the INCOMxxx.ZIP file and move the PIHSINC.xxx file to their library of PIHS data.

The 22 data sets that are generated by the programs listed below are finally brought together into a data set called PIHSINC.xxx (where xxx is DAT for ASCII, DTA for Stata, and SSP for SAS portable) using the program INCOME.SAS. This program adds up the income earned by the household from various sources into a variable called HHINCOME. Income components from various activities (agriculture, family enterprises, wages, and other sources of income) are also stored separately.

Note however that in this measure of household income, if households report negative incomes from either agriculture or family enterprise activities, the income from these sub-components is set to zero when calculating HHINCOME. The preliminary estimate of HHINCOME can clearly be improved upon. The data contain outliers, many of which can probably be corrected on closer inspection. Some have already been detected; however, there are probably many others that still remain, especially in the family enterprise section.

Dataset	PIHS files	Level of observation	Name of variable	Description / comments
Housing	F02B	Household	Hrent	For all households excluding renters. Either reported imputed rent used, or rent imputation based on hedonic regression.
W_agri	F05A1-2	Individual	W_agri	See notes in program regarding bonding payment and value of in-kind payments.
W_nagri	F05B1-5	Individual	W_nagri	Some individual fixes, and imputations based on averages for the profession. See notes in program for details.
Pension	F05C	Individual	Pension	Sum of pension and social security
Abroad	F05D	Individual	Abroad	Sum of remittances in cash and in-kind
Timeuse	F06C	Individual	Womeninc	Earnings for past 30 days x 12
Agr_inc1	F09B1-3	Household	Valcrop1	Sum of Rabi crops. See notes in program for details.

Dataset	PIHS files	Level of observation	Name of variable	Description / comments
Agr_inc2	F09B4-6	Household	Valcrop2	Sum of Kharif crops.
Agr_inc3	F09B7-8	Household	Valfruit	Sum of Orchard crops.
Sug_cane	F09B9	Household	Valcane	Value of sugarcane produced. See notes in the program regarding assumptions made.
Agrinput	F09DB F09DE F09DH F09DI	Household	Agrinput	Expenditure on agriculture inputs such as seeds fertilizer, insecticides, and other such inputs.
Agrrent1	F09DM	Household	Agrrent1	Income from renting farming machinery in past 12 months, net of maintenance and operating expenses.
Agrrent2	F09E1 F09E2	Household	Agrrent2	Income from selling water, hiring out animals, tractor and thresher rentals etc., net of expenses.
Lvstock	F09F2 F09F3	Household	Lvstock	Value of animals sold minus expenditure on inputs
Labor	F09G2 F09G3	Household	Labor	See notes in program

Dataset	PIHS files	Level of observation	Name of variable	Description / comments
Process	F09H2	Household	Process	Sum of earnings minus expenditures
Dairy	F09H3	Household	Dairy	Some high values. See program for details.
Entprise	F10A2, F10B F10C1, F10D	Household	Entprise	Outliers replaced with values from 95th percentile
Credit	F15C2	Household	Interest	See notes in program
Invments	F15DM	Household	Invments	Property rents received plus income from agr. land Note however that no households report income from agri. land (See note in program)
Remitin	F16B2	Household	H_remit O_remit	By household members and others. See note regarding this in program.
Otherinc	F17	Household	Otherinc	Sum of all categories listed in section 17.

APPENDIX 6: NOTES ON THE PIHS EXPENDITURE AGGREGATES:

Included with the raw data is a file containing an aggregated expenditure variable and all of the program files (as well as the supplementary data files) used to construct it. The consultant who constructed this aggregate has agreed to allow the distribution of this information with the understanding that the descriptions given in this document and in the program files are the only documentation that will be provided. The aggregated expenditure file can be found in a compressed file called EXPENxxx.ZIP, where xxx is DAT, DTA, or SSP, depending on whether the user has requested to receive ASCII, Stata, or SAS files. Decompressing this file results in two files[6]--the expenditure data file (PIHSEXPN.xxx) and another compressed file (BACKGRND.ZIP) which archives a document containing all of the programs and another compressed file containing all the supplementary data files (SUPPLxxx.ZIP). The supplementary data files in SUPPLxxx..ZIP are in ASCII, Stata, or SAS, depending on the requested format. The document containing all of the programs as well as some comments describing them are supplied both in WordPerfect 5.1 format and ASCII. (The two files are called PIHSEXPN.WP5 and PIHSEXPN.ASC.)

Any manipulation of the data requires that assumptions be made and, to the extent possible, those assumptions are explained below and in the document found in BACKGRND.ZIP. Given the complexity and detail involved in the different expenditure modules, it is possible to construct an expenditure aggregate in different ways. Any researcher not satisfied with the assumptions made in this expenditure aggregate should build their own estimate from the household data or alter the program files provided with this expenditure aggregate. Researchers who are satisfied with the assumptions made, may choose to simply use the PIHSEXPN.xxx file and not bother opening the BACKGRND.ZIP file. The following describes part of the process in constructing the expenditure data file (PIHSEXPN.xxx) and its contents.

---

[6]In the case of the ASCII format it will result in three files where the additional file is the dictionary for the data file.

General notes on PIHS consumption data

The PIHS survey was based on two visits, approximately a fortnight apart, with the bulk of the consumption data being collected in the second visit. The reference period is either a typical month in the year or the time period since the first visit, and sometimes both, depending on the type of good. Where appropriate, information on non-purchased as well as purchased consumption is requested. In addition, the PIHS asks specifically about payment-in-kind.

With the PIHS, there are choices to be made concerning the determination of the component variables. The first choice where choices need to be made is the reference period. For all purchased food-stuffs and for some personal use items, respondents were asked about both their consumption since the last visit and their consumption in a typical month. In every case, we use the monthly information. Second, for this reference period, for food, value information only is provided for purchases, quantity information only is provided for gifts-in-kind, and payment-in-kind and both value and quantity information are provided for self-produced consumption. We use reported value information where available (ie., for purchases and self-produced consumption). Where this is unavailable, we use prices to convert quantities into values. The prices are obtained from a variety of sources. Where possible, they are taken from the community questionnaire which contained a price survey. Other prices are obtained from the unit-values (either those of the household itself if available, or the average of those of the PSU if available, or the average of those of the province if available or, as a last resort, the nationwide average) obtained from information on purchases since the first visit.

The second area in which choices need to be made is the value of consumption. There are two possible methods are for the valuation of durable consumption. Information on time of purchase, value at time of purchase and current value are all requested in the PIHS so that the depreciation of all durables can be calculated and used to estimate an imputed income flow from them. However, one can also use the information provided on time of purchase and value at time of purchase to calculate actual expenditures on durables over the last twelve months, which is the information sought in the HIES. While both measures have been calculated using PIHS data[7],

---

[7] A purchase of a durable within the last year was assumed to have taken place if the number of years of acquisition (which is recorded as an integer) was non-missing and less than one (ie., zero). Jewelery was excluded from the income-flow calculations.

the imputed rental value of the durable goods is the measure included in the total expenditure variable.

Finally, choices must be made regarding housing expenditures. Respondents were asked to estimate their imputed rent on their housing. They were also asked to reveal information about their housing conditions. This information allows an expected rent to be calculated by the regression of reported rent on housing characteristics. The total expenditure variable uses expected rent rather than reported rent. There are two advantages to working with expected or hedonic rather than reported rent. First, for one-tenth of the survey the reported rental value of the accommodation is missing. Second, the expected rent value will reduce the noise associated with people's misunderstandings of the value of their accommodation. But there are countervailing advantages to working with reported rental values. First, since it is impossible to capture all relevant housing information, the expected rent may fail to reflect house-specific features. Second, the HIES rental values are reported values. Using reported values for the PIHS will aid in comparability. The processed data set uses only one reference period, but contains variables which allow for both methods of calculation of durable and rental expenditure. See the notes to the table below for details.

Cleaning of data

The data cleaning is carried out in the various macros dealing with the different categories of consumption (and household size). See these macros in the PIHSEXPN.WP5 file for details. When component data sets were merged, only observations with a non-missing household code were accepted. This condition resulted in a data set of 4,799 households.

The following conditions were applied to all observations in PIHSEXPN:

(a)   The food share (VEXP1000/VEXP0000) had to be between .5 and 90%;

(b)   Household size had to be greater than zero; and

(c)   The raising factor had to be greater than zero.

Households which satisfy all three of these conditions have a value of unity for the variable GOOD. There are 4,745 such households.

Further cleaning may be required for particular exercises. For example, one household - 221201611 - has a believable value of consumption of vegetables - 100 Rs per month - but an unbelievable quantity - 20,020 kg per month.

Creation of PIHS expenditure data set

The construction of the PIHS set was done in two stages. In the first stage, the various components of aggregate expenditure were identified and computed and brought together into one large data set. In the second stage, the variables were re-coded to give a coding compatible with that in the HIES previously carried out in Pakistan. The data set resulting from the two stages is the PIHSEXPN.xxx data file.

The programs in the first stage used the data included with the PIHS household data files. There were, however, some exceptions to this general rule where data from outside sources was used. These data files are found in the SUPPLxxx.ZIP file, which is archived in the BACKGRND.ZIP file. The exceptions are as follows:

(a)   The program constructing aggregate energy expenditures was missing. Hence for this category, an ASCII data set (ENERGY.OUT) is required containing the orginally calculated aggregate energy expenditure figures.

(b)   Food prices are required to convert quantities received in kind to values received in kind, and to convert values of purchases into quantities of purchases. Computation of these prices is a complex task. Sources included: (i) prices in the community surveys; (ii) average unit values from quantities and values over the last two weeks; and (iii) information from the FBS.   The data set PRICTOT was used to measure the required food prices.

(c)   Special care was taken with food codes 333-335. In the Urdu questionnaire, for these three codes the question which asks for the value of purchased consumption in a typical month in blacked out. Interviewers were instructed, nevertheless, to ask the question. Some did. Others didn't. A large number of missing values resulted. For these, regressions were run. The regressions used education of the head of the household among other variables. Rather than reconstruct this variable from scratch, the constructed data set ED, which contains this variable, is used.

(d)   Health consumption data from the health module (not the consumption module) was extracted by a consultant for the World Bank and is contained in the constructed data set, HEALTH

(e)   Sampling weights are contained in a separate data set, WEIGHTS.

In summary, one needs the following raw data sets for the first stage: F01A, F02A, F02B, F02C, F03B2, F05B2, F05B3, F11A, F11B, F11C, F12A, F12B2, F16A2.   To construct PIHSEXPN.xxx, one also needs the following supplementary data sets contained in SUPPLxxx.ZIP: WEIGHTS,PRICTOT, ED, ENERGY.OUT, HEALTH.EXP.

Using the program PCREATE and the macros referred to therein, these data sets create a preliminary data set called PIHSEXP. This data set is not available for distribution. (All quantity and monetary values in the data set resulting from the first stage are annual, while the values in the data set resulting from the second stage are monthly values.)

The second stage uses the PIHSEXP preliminary data file and applies the program RENAME, which does the following:

(a)   Renames variables to make them consistent with HIES conventions and to bring out clearly the structure of the data set;

(b)   Labels all variables;

(c)   Cleans data set (very briefly); and

(d)   Puts all variables in monthly terms.

The data set resulting from this stage, PIHSEXPN, contains a measure of total monthly expenditure as well as several subcomponents. The tables below list the contents and describes the naming conventions used in PIHSEXPN.

CONTENTS OF PIHSEXPN AND NOTES TO THE DATA

```
                                                      F                  P       U          S
         N           U          S                     R                  R       S          U          H
         A           N          O                     O                  E       A    C     B    C     E
         M           I    T     U          M    G          I    S    C   G    A    G    A    A
O        E           T    Y     U          O    I    P     D    E    O   A    L    P    T    D
B        9           9    P     R          W    F    U     9    N    D   G    9    9    C    9
S        1           1    E     E          N    T    R     1    T    E   E    1    1    H    1

  1  all expenditure            .     .        .      Yes  Yes  Yes  0000  0    .     .        .     0    1   1
  2  food                       .     .        S12F   Yes  Yes  Yes  1000  0    .     .        .   100    2   1
  3  cereals                    .     CEREALS  S12F   Yes  Yes  Yes  1010  0    .     .        .   101    3   1
  4  wheat (grain)              kg    CEREALS  S12F   Yes  Yes  Yes  1011  0   301   1.00   3490   101    4   0
  5  wheat (maida)              kg    CEREALS  S12F   No   Yes  Yes  1012  1   302   1.00   3490   101    5   0
  6  fine rice (Basmati)        kg    CEREALS  S12F   Yes  Yes  Yes  1013  1   305   1.00   3640   101    6   0
  7  coarse rice                kg    CEREALS  S12F   Yes  Yes  Yes  1014  1   306   1.00   3640   101    7   0
  8  maize (flour or grain)     kg    CEREALS  S12F   Yes  Yes  Yes  1015  1   303   1.00   3490   101    8   0
  9  jawar/Bajra                kg    CEREALS  S12F   Yes  Yes  Yes  1016  1   304   1.00   3500   101    9   0
 10  other grains or cereals    kg    CEREALS  S12F   Yes  Yes  Yes  1019  1   307   1.00   3500   101   10   0
 11  baked products             .     BAKED    S12F   No   Yes  Yes  1020  0    .     .        .   102   11   1
 12  biscuits/cakes             kg    BAKED    S12F   No   Yes  Yes  1021  1   334   1.00   1018   102   12   0
 13  breads (bhapati nun)       kg    BAKED    S12F   No   Yes  Yes  1022  1   332   1.00    134   102   13   0
 14  fried items (samosa  pakora) kg  BAKED    S12F   No   Yes  Yes  1029  1   333   1.00    260   102   14   0
 15  pulses                     .     PULSES   S12F   Yes  Yes  Yes  1030  0    .     .        .   103   15   1
 16  gram                       kg    PULSES   S12F   Yes  Yes  Yes  1031  1   308   1.00   3750   103   16   0
 17  dal                        kg    PULSES   S12F   Yes  Yes  Yes  1039  1   309   1.00   3610   103   17   0
 18  milks                      .     MILKS    S12F   Yes  Yes  Yes  1040  0    .     .        .   104   18   1
 19  fresh milk                 lt    MILKS    S12F   Yes  Yes  Yes  1041  1   313   1.00    660   104   19   0
 20  milk powder                kg    MILKS    S12F   No   Yes  Yes  1043  1   315   1.00   5010   104   20   0
 21  baby formula               kg    MILKS    S12F   No   Yes  Yes  1044  1   316   1.00   3570   104   21   0
 22  ghee+desi ghee             kg    MILKS    S12F   Yes  Yes  Yes  1045  1   312   1.00   9000   104   22   0
 23  yoghurt                    kg    MILKS    S12F   Yes  Yes  Yes  1047  1   314   1.00    700   104   23   0
 24  oils                       kg    OILS     S12F   Yes  Yes  Yes  1050  0   311   1.00   9000   105   24   1
 25  meats                      .     MEATS    S12F   Yes  Yes  Yes  1060  0    .     .        .   106   25   1
 26  mutton lamb goat           kg    MEATS    S12F   Yes  Yes  Yes  1061  1   319   0.75   1640   106   26   0
 27  beef buffaloe              kg    MEATS    S12F   Yes  Yes  Yes  1062  1   320   0.85   2120   106   27   0
 28  fish                       kg    MEATS    S12F   Yes  Yes  Yes  1063  1   323   0.95   1120   106   28   0
 29  poultry and eggs           .     POULTEGG S12F   Yes  Yes  Yes  1070  0    .     .        .   107   29   1
 30  chicken                    kg    POULTEGG S12F   Yes  Yes  Yes  1071  1   321   0.60   1850   107   30   0
 31  egg                        No.   POULTEGG S12F   Yes  Yes  Yes  1072  1   322   0.90     75   107   31   0
 32  fruit                      .     FRUIT    S12F   Yes  Yes  Yes  1080  0    .     .        .   108   32   1
 33  bananas                    .     FRUIT    S12F   Yes  Yes  Yes  1081  1   326   0.65    157   108   33   0
 34  citrus fruits              .     FRUIT    S12F   Yes  Yes  Yes  1082  1   327   0.75     68   108   34   0
 35  mango                      kg    FRUIT    S12F   Yes  Yes  Yes  1083  1   328   0.70    640   108   35   0
 36  melon                      .     FRUIT    S12F   Yes  Yes  Yes  1085  1   325   0.65    210   108   36   0
 37  other                      .     FRUIT    S12F   Yes  Yes  Yes  1089  1   329   1.00    480   108   38   0
 38  vegetables                 kg    VEG      S12F   Yes  Yes  Yes  1090  1   324   1.00    400   109   39   1
 39  spices (and condiments)    kg    SPICES   S12F   Yes  Yes  Yes  1100  1   335   1.00   3360   110   40   1
 40  sugar                      .     SUGAR    S12F   No   Yes  Yes  1110  0    .     .        .   111   41   1
```

OBS	NAME91	UNIT91	TYPE	SOURCE	FROMOWN	GIFT	PUR	ID91	PRESENT	CODE91	USAGE91	CAL91	SUBGP91	CATCH	HEAD91
41	refined sugar	kg	SUGAR	S12F	No	Yes	Yes	1111	1	317	1	3910	111	42	0
42	desi sugar (gur)	kg	SUGAR	S12F	Yes	Yes	Yes	1112	1	318	1	3710	111	43	0
43	tea and coffee	.	TEACOFF	S12F	No	Yes	Yes	1120	0	.	.	.	112	44	1
44	tea	kg	TEACOFF	S12F	No	Yes	Yes	1121	1	336	1	2900	112	45	0
45	coffee	kg	TEACOFF	S12F	No	Yes	Yes	1122	1	337	1	1340	112	46	0
46	bottled drinks (cola squash etc)	.	OTHER	S12F	No	Yes	Yes	1130	1	331	1	100	113	47	1
47	tobacco cigarettes naswar pan	.	TOBACCO	S11AF	No	Yes	Yes	1150	1	104	.	.	115	48	1
48	hh food/clothing recd from emper	.	TFCLTHYR	S5Bq21M+F.	No	No	No	1160	1	.	.	.	116	81	1
49	canned food	.	OTHER	S12F	No	Yes	Yes	1180	1	330	1	1200	118	37	1
50	other foods	.	OTHER	S12F	Yes	Yes	Yes	1190	0	.	.	.	119	49	1
51	ground nuts	.	OTHER	S12	Yes	Yes	Yes	1191	1	310	1	5110	119	50	0
52	miscellaneous food expenses	.	OTHER	S12	Yes	Yes	Yes	1199	1	338	.	.	119	51	0
53	fuel and lighting	.	.	.	No	No	No	2000	0	.	.	.	200	52	1
54	expenditure on energy	.	ENEREXP	S7M	No	No	Yes	2001	1	.	.	.	200	53	0
55	kerosene matches and candles	.	S11A	S11A	No	Yes	Yes	2113	1	106	.	.	211	54	0
56	personal use items	.	.	.	No	No	No	3000	0	.	.	.	300	55	1
57	clothing	.	.	.	No	No	No	3210	0	.	.	.	321	56	1
58	children clothing and material	.	S11B	S11B	No	Yes	Yes	3211	1	120	.	.	321	57	0
59	adult clothing and material	.	S11B	S11B	No	Yes	Yes	3212	1	121	.	.	321	58	0
60	Footwear	.	.	.	No	No	No	3220	0	.	.	.	322	59	1
61	children footwear	.	S11B	S11B	No	Yes	Yes	3221	1	122	.	.	322	60	0
62	adult footwear	.	S11B	S11B	No	Yes	Yes	3222	1	123	.	.	322	61	0
63	other personal effects	.	S11B	S11B	No	Yes	Yes	3230	1	124	.	.	323	62	1
64	stitching or repair of wearing apparel	.	S11B	S11B	No	Yes	Yes	3240	1	125	.	.	324	63	1
65	jewelry	.	DURABLES	S11CF	No	Yes	Yes	3312	1	204	.	.	331	126	0
66	household textiles	.	S11B	S11B	No	Yes	Yes	3320	1	127	.	.	332	64	1
67	housing	.	.	.	No	No	No	4000	0	.	.	.	400	65	1
68	rent and housing expenitures	.	.	.	No	No	No	4210	0	.	.	.	421	66	1
69	actual rent	.	RENT_ACT	S2CM	No	No	Yes	4211	1	.	.	.	421	67	0
70	imputed rent	.	RENT_IMP	S2CM	No	No	Yes	4213	1	.	.	.	421	68	0
71	repair and maintenance of house	.	S11B	S11B	No	Yes	Yes	4214	1	133	.	.	421	69	0
72	housing and property taxes	.	S11B	S11B	No	Yes	Yes	4215	1	135	.	.	421	70	0
73	annual garbage disposal expenditure	.	GARBYR	S2BM	No	No	Yes	4216	1	.	.	.	421	71	0
74	annual water expenditure	.	WATERYR	S2BM	No	No	Yes	4217	1	.	.	.	421	72	0
75	annual utility repairs	.	UTILREP	S2BF	No	No	Yes	4219	1	.	.	.	421	73	0
76	repair and servicing of hh effects	.	S11B	S11B	No	Yes	Yes	4230	1	130	.	.	423	74	1
77	other household effects	.	S11B	S11B	No	Yes	Yes	4240	1	129	.	.	424	75	1
78	kitchen equipment incl crockery	.	S11B	S11B	No	Yes	Yes	4320	1	126	.	.	432	76	1
79	furniture and fittings	.	S11B	S11B	No	Yes	Yes	4330	1	128	.	.	433	77	1
80	other durable housing expenditure	.	.	.	No	No	No	4390	0	.	.	.	439	78	1

OBS	NAME91	UNIT91	TYPE	SOURCE	FROMOWN	GIFT	PUR	ISD91	PRESENT	CODE	USAGE91	CAL91	SUBGAP91	CATCH	HEAD91
81	home improvements and additions	.	S11B	S11B	No	Yes	Yes	4398	1	134	.	.	439	79	0
82	land/buildings for residence/investment	.	S11B	S11B	No	Yes	Yes	4399	1	143	.	.	439	80	0
83	miscellaneous	.	.	.	No	No	No	5000	0	.	.	.	500	82	1
84	toiletries	.	.	.	No	No	No	5110	0	.	.	.	511	83	1
85	commercial or handmade soap	.	S11A	S11A	No	Yes	Yes	5111	1	101	.	.	511	84	0
86	oth pers care (cosmtcs soap cmbs etc)	.	S11A	S11A	No	Yes	Yes	5119	1	101	.	.	511	85	0
87	personal services (eg haircut shoeshine)	.	S11B	S11B	No	Yes	Yes	5120	1	139	.	.	512	86	1
88	recreation and travel	.	.	.	No	No	No	5130	0	.	.	.	513	87	1
89	newspapers books and other entertainment	.	S11A	S11A	No	Yes	Yes	5131	1	105	.	.	513	88	0
90	recreation personal travel lodging	.	S11B	S11B	No	Yes	Yes	5133	1	138	.	.	513	89	0
91	meals eaten outside the house	.	MEALSOUT	S11BF	No	Yes	Yes	5134	1	107	.	.	513	90	0
92	personal transport expenses	.	.	.	No	No	No	5140	0	.	.	.	514	91	1
93	gas motor oil for personal transport	.	S11A	S11A	No	Yes	Yes	5141	1	103	.	.	514	92	0
94	repair/service of vehicles excl gas+oil	.	S11B	S11B	No	Yes	Yes	5142	1	131	.	.	514	93	0
95	public transport incl rickshaws+taxis	.	S11B	S11B	No	Yes	Yes	5144	1	132	.	.	514	94	0
96	transport subsidy from employer	.	TTRNSPYR	S5Bq23M+F	No	Yes	No	5149	1	.	.	.	514	133	0
97	misc frequently incurred expenditure	.	.	.	No	No	No	5190	0	.	.	.	519	95	1
98	wages to servants gardeners etc	.	S11A	S11A	No	Yes	Yes	5191	1	108	.	.	519	96	0
99	postal articles telegram telephone	.	S11B	S11B	No	Yes	Yes	5192	1	142	.	.	519	97	0
100	annual telephone expenditure	.	TELEPHYR	S2BM	No	No	Yes	5193	1	.	.	.	519	98	0
101	health expenses	.	.	.	No	No	No	5210	0	.	.	.	521	99	1
102	non-diah. health services	.	HEALTH	HTH MOD	No	No	Yes	5213	1	.	.	.	521	100	0
103	diah. health services	.	HEALTHD	HTH MOD	No	No	Yes	5217	1	.	.	.	521	101	0
104	education	.	.	.	No	No	No	5240	0	.	.	.	524	102	1
105	adminn/regn/tuition	.	EXPADM	S3F	No	No	Yes	5241	1	.	.	.	524	105	0
106	uniforms	.	EXPUNF	S3F	No	No	Yes	5242	1	.	.	.	524	106	0
107	books for education	.	EXPBKS	S3F	No	No	Yes	5243	1	.	.	.	524	107	0
108	transport for education	.	EXPTR	S3F	No	No	Yes	5244	1	.	.	.	524	108	0
109	private tuition	.	EXPTUT	S3F	No	No	Yes	5245	1	.	.	.	524	109	0
110	exam fees	.	EXPEXAM	S3F	No	No	Yes	5246	1	.	.	.	524	110	0
111	other education expenditure	.	EXPOTHS	S3F	No	No	Yes	5247	1	.	.	.	524	111	0
112	unspecified education expenditure	.	EXPUNSPC	S3F	No	No	Yes	5248	1	.	.	.	524	112	0
113	received by household in scholarships	.	HHTVSCH	S3Fq19	No	No	Yes	5249	1	.	.	.	524	103	0
114	help received for educational expenses	.	HHTVOPTU	S3Fq21	No	No	Yes	5250	1	.	.	.	525	104	1
115	ed/profnl services repted in consn sctn	.	S11B	S11B	No	Yes	Yes	5251	1	140	.	.	525	113	0
116	stationery books (non-education-related)	.	S11B	S11B	No	Yes	Yes	5260	1	141	.	.	526	114	1
117	misc infrequently incurred expenditure	.	.	.	No	No	No	5290	0	.	.	.	529	115	1
118	cash losses	.	S11B	S11B	No	Yes	Yes	5291	1	148	.	.	529	116	0
119	marriages births and other ceremonies	.	S11B	S11B	No	Yes	Yes	5292	1	145	.	.	529	117	0
120	funerals and related death expenses	.	S11B	S11B	No	Yes	Yes	5293	1	144	.	.	529	118	0

OBS	NAME91	UNIT91	TYPE	SOURCE	FROMOWN	GIFT	PUR	ID91	PRESENT	USACODE	CLP91	CAGP91	SUBCAGH	CATCH91	HEAD91
121	legal expenses	.	S11B	S11B	No	Yes	Yes	5294	1	147	.	.	529	119	0
122	remittances to household members	.	TSENTMEM	S16AMq10	No	No	No	5295	1	.	.	.	529	120	0
123	dowry	.	S11B	S11B	No	Yes	Yes	5299	1	146	.	.	529	121	0
124	miscellaneous durable epxenses	.	DURABLES	S11CF	No	No	No	5300	0	.	.	.	530	122	1
125	radio	.	DURABLES	S11CF	No	Yes	Yes	5321	1	201	.	.	532	123	0
126	gramophone/phonograph/tape-player	.	DURABLES	S11CF	No	Yes	Yes	5326	1	202	.	.	532	124	0
127	camera	.	DURABLES	S11CF	No	Yes	Yes	5331	1	203	.	.	533	125	0
128	guns	.	DURABLES	S11CF	No	Yes	Yes	5332	1	205	.	.	533	127	0
129	bicycle	.	DURABLES	S11CF	No	Yes	Yes	5341	1	206	.	.	534	128	0
130	motorcycle/scooter	.	DURABLES	S11CF	No	Yes	Yes	5342	1	207	.	.	534	129	0
131	automobile/truck	.	DURABLES	S11CF	No	Yes	Yes	5343	1	209	.	.	534	131	0
132	motor rickshaw	.	DURABLES	S11CF	No	Yes	Yes	5349	1	208	.	.	534	130	0
133	other durables	.	DURABLES	S11CF	No	Yes	Yes	5399	1	210	.	.	539	132	0

Notes: The table gives the consumption information contained in the SAS data set PIHSEXPN. The prefix for all the consumption value codes is VEXP (value of expenditure). The code of the HIES variable corresponding most closely to the relevant PIHS variable has been chosen, where one exists. However, the PIHS variables are defined by their labels, not by the labels of the corresponding HIES codes in Tables A1 or A2. Sometimes the correspondence is at best weak. The column 'Source' gives the source of the PIHS variable. Here, for example, 'S11aF' means that the variable comes from **S**ection **11a** of the questionnaire, which is one of the **f**emale sections. The 'original code' gives the code of the variable in the questionnaire if one exists. If the source column is empty, the variable is an aggregate, missing from the raw data. There are various expenditure variables in the data set not reported in the table:

(a) V2XP4213 is regressed imputed rent. (VEXP4213, given in the table, is reported imputed rent - except where this is missing, in which case it is set equal to regressed rent (ie., VEXP4213=V2XP4213).)

(b) V2XP5300 is the estimated income flow derived from miscellaneous durables (5321, 5326, 5331, 5332, 5341, 5342, 5343, 5349, 5399 - all appearing in the data set, prefixed V2XP, and with codes corresponding to the VEXP variables) based on their calculated depreciation.

(c) Consumption data is available in the PIHS on that amount of expenditure due to purchases, that amount due to gifts-in-kind and payment-in-kind, and that amount imputed from self-production. Expenditure in the form of gifts-in-kind and payment-in-kind is prefixed by a GV. Expenditure in the form of purchases is prefixed by a PV. Expenditure in the form of self-production is prefixed by an FV. Codes are as in the table above. Thus FV1011 is the value of wheat self-produced. If quantities are available, they are prefixed by GQ, PQ or FQ. The prefix Q gives the sum of these three. (Quantities have been calculated for all foodstuffs except the miscellaneous category.)

(d) V2XP5295 is all remittances, whereas VEXP5295 is remittances to hosuehold members.

(e) V2XP1600, V3XP1600 AND V4XP1600 are, respectively, free or subsidized housing from the employer, free or subsidized transport from the employer, and free or subsidized other payments (excluding wages, bonuses, the above-mentioned benefits and food and clothing). All refer to primary off-farm employment: as for VEXP1600 in the table above, see Section 5B of the female questionnaire.

(f) V2XP5211 P2V5211 G2V5211 are total, purchased and gifted health expenditure corresponding to item 136 in the consumption module; likewise 5213 corresponds to 137; and 5210 is the total of these two (data from health module used in preference).

All consumption figures in PIHSEXPN - values and quantities - are given in monthly terms.

PIHSEXPN also contains some non-consumption variables. PROVINCE indicates the province, URBRURAL urban or rural, GOOD whether the variable has passed the cleaning test (see text). WEIGHT is the raising factor and HHCODE the household code. Then there are four household size variables. HHSIZE counts all household members. HHSIZEW weights membership by time of residence in the year. HHSIZE2 gives a definition of household size similar to that used by the HIES. HHSIZE3 counts as household members only those who are present throughout the year.

Note that no distinction is made for any of the variables between missing values and zeroes.

CONTENTS & DESCRIPTIVE STATISTICS OF PIHSEXPN

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
HHCODE	Household code	4745	175262503	4315105932	110100101	420403323
HHSIZE	Household size	4745	7.3478925	190.9801271	1.0000000	40.0000000
HHSIZEW	Household size weighting by residence	4745	7.0750886	186.8547644	0.7500000	39.0000000
HHSIZE2	Household size using HIES defn	4745	7.1685476	187.0116511	1.0000000	40.0000000
HHSIZE3	Hh size counting only permt residents	4704	6.8545837	185.8705465	1.0000000	38.0000000
PROVINCE	Province P=1 S=2 N=3 B=4	4745	1.5633481	43.1494554	1.0000000	4.0000000
URBRURAL	Urban=1 Rural=2	4745	1.6986171	24.0271035	1.0000000	2.0000000
VEXP0000	all expenditure	4745	4092.36	263616.73	213.8333333	242513.67
VEXP1000	food	4745	1795.57	129437.80	66.0833333	183616.33
VEXP1010	cereals	4745	372.6394293	28991.52	0	28520.00
PV1011	wheat (grain)	4745	55.5017427	6246.70	0	2100.00
GV1011	wheat (grain)	4745	0.3398713	417.7001179	0	395.8333333
FV1011	wheat (grain)	4745	103.6768626	26469.28	0	28500.00
PQ1011	wheat (grain)	4745	17.1399495	1896.59	0	500.9174312
GQ1011	wheat (grain)	4745	0.0888123	92.2051149	0	83.3333333
FQ1011	wheat (grain)	4745	23.9789316	2877.31	0	1350.00
VEXP1011	wheat (grain)	4745	159.5184766	27060.74	0	28500.00
VQ1011	wheat (grain)	4745	41.2076934	3270.55	0	1350.00
PV1012	wheat (maida)	4745	119.2444272	8331.93	0	3160.00
GV1012	wheat (maida)	4745	0.2370151	409.4305398	0	385.4166667
VEXP1012	wheat (maida)	4745	119.4814423	8351.31	0	3160.00
PQ1012	wheat (maida)	4745	29.9862982	2067.63	0	842.6666667
GQ1012	wheat (maida)	4745	0.0527369	89.3921493	0	83.3333333
VQ1012	wheat (maida)	4745	30.0390350	2071.34	0	842.6666667
PV1013	fine rice (Basmati)	4745	31.6955477	3854.57	0	1300.00
GV1013	fine rice (Basmati)	4745	0.4167456	608.1993997	0	500.0000000
FV1013	fine rice (Basmati)	4745	11.9844682	3421.36	0	3000.00
PQ1013	fine rice (Basmati)	4745	2.8914032	342.1273081	0	100.0000000
GQ1013	fine rice (Basmati)	4745	0.0414545	61.4290478	0	50.0000000
FQ1013	fine rice (Basmati)	4745	1.2448015	351.0943209	0	300.0000000
VEXP1013	fine rice (Basmati)	4745	44.0967615	5115.13	0	3000.00
VQ1013	fine rice (Basmati)	4745	4.1776592	487.5638591	0	300.0000000
PV1014	coarse rice	4745	24.8902628	2923.36	0	840.0000000
GV1014	coarse rice	4745	0.1060456	187.6720900	0	400.0000000
FV1014	coarse rice	4745	7.3193216	3225.97	0	2400.00
PQ1014	coarse rice	4745	4.2609668	510.7001385	0	186.6666667
GQ1014	coarse rice	4745	0.0165128	30.0828666	0	66.6666667
FQ1014	coarse rice	4745	1.5817006	929.6900677	0	1001.00
VEXP1014	coarse rice	4745	32.3156300	4294.95	0	2400.00
VQ1014	coarse rice	4745	5.8591802	1050.66	0	1001.00
PV1015	maize (flour or grain)	4745	2.8271653	997.4432848	0	800.0000000
GV1015	maize (flour or grain)	4745	0.1266863	153.3962932	0	115.6250000
FV1015	maize (flour or grain)	4745	5.9583842	1784.54	0	500.0000000
PQ1015	maize (flour or grain)	4745	0.6290378	220.5740450	0	177.7777778
GQ1015	maize (flour or grain)	4745	0.0284352	34.4253376	0	25.0000000
FQ1015	maize (flour or grain)	4745	1.3733848	409.5309485	0	120.0000000
VEXP1015	maize (flour or grain)	4745	8.9122358	2268.88	0	858.3333333
VQ1015	maize (flour or grain)	4745	2.0308577	510.5480139	0	177.7777778
PV1016	jawar/Bajra	4745	0.7275892	545.3343329	0	400.0000000
GV1016	jawar/Bajra	4745	0.0060761	13.3413468	0	12.7777778
FV1016	jawar/Bajra	4745	1.2882800	1269.09	0	1500.00
PQ1016	jawar/Bajra	4745	0.1348333	98.6787872	0	80.0000000
GQ1016	jawar/Bajra	4745	0.000843464	1.7594129	0	1.6666667
FQ1016	jawar/Bajra	4745	0.2652915	208.1652013	0	150.0000000
VEXP1016	jawar/Bajra	4745	2.0219453	1384.41	0	1500.00
VQ1016	jawar/Bajra	4745	0.4009682	230.9605411	0	150.0000000
PV1019	other grains or cereals	4745	5.6973217	904.6499155	0	500.0000000
GV1019	other grains or cereals	4745	0.0234934	91.8272869	0	166.6666667
FV1019	other grains or cereals	4745	0.5721227	803.6876830	0	900.0000000
PQ1019	other grains or cereals	4745	0.7301351	125.3657411	0	43.9560440
GQ1019	other grains or cereals	4745	0.0025899	10.2441732	0	16.6666667
FQ1019	other grains or cereals	4745	0.0760680	96.1256498	0	90.0000000

```
Variable  Label                              N        Mean      Std Dev    Minimum     Maximum
----------------------------------------------------------------------------------------------

VEXP1019  other grains or cereals          4745    6.2929378     1356.34         0      1200.00
VQ1019    other grains or cereals          4745    0.8087930  169.1737787        0    108.7500000
VEXP1020  baked products                   4745   22.7995280     2379.87         0    948.2916667
PV1021    biscuits/cakes                   4745   12.2057916     1065.56         0    600.0000000
GV1021    biscuits/cakes                   4745    0.0250665  107.0558235        0    171.8750000
VEXP1021  biscuits/cakes                   4745   12.2308581     1070.96         0    600.0000000
PQ1021    biscuits/cakes                   4745    1.6536595  149.5701969        0     75.0000000
GQ1021    biscuits/cakes                   4745    0.0030537   12.9816203        0     20.8333333
VQ1021    biscuits/cakes                   4745    1.6567132  150.1280279        0     75.0000000
PV1022    breads (bhapati nun)             4745    4.5885335     1631.39         0    900.0000000
GV1022    breads (bhapati nun)             4745    0.0058501   22.6212757        0     41.6666667
VEXP1022  breads (bhapati nun)             4745    4.5943836     1632.05         0    902.5000000
PQ1022    breads (bhapati nun)             4745    3.5745571     1403.70         0    900.0000000
GQ1022    breads (bhapati nun)             4745    0.0058501   22.6212757        0     41.6666667
VQ1022    breads (bhapati nun)             4745    3.5804071     1404.49         0    902.5000000
PV1029    fried items (samosa  pakora)     4745    5.9727070  619.3695271        0    200.0000000
GV1029    fried items (samosa  pakora)     4745    0.0015794    4.1654526        0      8.3333333
VEXP1029  fried items (samosa  pakora)     4745    5.9742864  619.3569746        0    200.0000000
PQ1029    fried items (samosa  pakora)     4745    5.8107841  636.0071833        0    204.0816327
GQ1029    fried items (samosa  pakora)     4745    0.0015845    4.1659338        0      8.3333333
VQ1029    fried items (samosa  pakora)     4745    5.8123686  635.9971614        0    204.0816327
VEXP1030  pulses                           4745   53.4355563     2692.56         0    760.0000000
PV1031    gram                             4745   10.2434196  828.3190903        0    450.0000000
GV1031    gram                             4745    0.1311233  308.8642626        0    322.3214291
FV1031    gram                             4745    0.7592244  437.4940857        0    360.0000000
PQ1031    gram                             4745    1.2162506  115.2309822        0    100.0000000
GQ1031    gram                             4745    0.0196222   45.4227658        0     41.6666667
FQ1031    gram                             4745    0.1216910  128.3169330        0    150.0000000
VEXP1031  gram                             4745   11.1337673  975.4563454        0    450.0000000
VQ1031    gram                             4745    1.3575637  177.3988426        0    150.0000000
PV1039    dal                              4745   40.6458135     2184.62         0    680.0000000
GV1039    dal                              4745    0.1922989  590.2456733        0    750.0000000
FV1039    dal                              4745    1.4636766  698.1747769        0    300.0000000
PQ1039    dal                              4745    2.2945754  121.9709898        0     37.7777778
GQ1039    dal                              4745    0.0110328   33.5690436        0     41.6666667
FQ1039    dal                              4745    0.1162003   63.1935791        0     50.0000000
VEXP1039  dal                              4745   42.3017890     2326.45         0    760.0000000
VQ1039    dal                              4745    2.4218085  138.6712641        0     50.7500000
VEXP1040  milks                            4745  457.6784242   105671.78         0    178378.00
PV1041    fresh milk                       4745  129.6585035     8467.36         0      1800.00
GV1041    fresh milk                       4745    0.5220960  519.8333594        0    375.0000000
FV1041    fresh milk                       4745  137.8752101    15154.71         0      5715.00
PQ1041    fresh milk                       4745   19.6130786     1245.60         0    300.0000000
GQ1041    fresh milk                       4745    0.0890908   89.1547116        0     75.0000000
FQ1041    fresh milk                       4745   30.3514466    13483.01         0     15000.00
VEXP1041  fresh milk                       4745  268.0558096    14815.80         0      5895.00
VQ1041    fresh milk                       4745   50.0536160    13448.66         0     15017.14
PV1043    milk powder                      4745    1.5299970     1028.82         0    760.0000000
GV1043    milk powder                      4745    0.0063987   16.3577734        0     17.5000000
VEXP1043  milk powder                      4745    1.5363957     1030.15         0    760.0000000
PQ1043    milk powder                      4745    0.0187917   12.5918255        0      8.8578089
GQ1043    milk powder                      4745  0.000091410    0.2336825        0      0.2500000
VQ1043    milk powder                      4745    0.0188831   12.6141525        0      8.8578089
PV1044    baby formula                     4745    1.1851056     1081.18         0    760.0000000
GV1044    baby formula                     4745  0.000142996    2.8465436        0     20.6666667
VEXP1044  baby formula                     4745    1.1852485     1081.28         0    760.0000000
PQ1044    baby formula                     4745    0.0070860    6.6558629        0      5.3333333
GQ1044    baby formula                     4745 5.7659597E-7    0.0114780        0      0.0833333
VQ1044    baby formula                     4745    0.0070866    6.6561086        0      5.3333333
PV1045    ghee+desi ghee                   4745   18.7037441     2482.49         0    840.0000000
GV1045    ghee+desi ghee                   4745    0.1034259   90.5336832        0     64.2857143
FV1045    ghee+desi ghee                   4745  102.8628956   102617.42         0    177000.00
PQ1045    ghee+desi ghee                   4745    0.5080988   79.8045126        0     22.0000000
GQ1045    ghee+desi ghee                   4745    0.0016145    1.3912156        0      1.0000000
FQ1045    ghee+desi ghee                   4745    1.5401010  214.5748750        0     80.0000000
VEXP1045  ghee+desi ghee                   4745  121.6700656   102610.87         0    177000.00
VQ1045    ghee+desi ghee                   4745    2.0498143  221.6832400        0     80.0000000
```

68

```
Variable  Label              N         Mean        Std Dev    Minimum        Maximum
----------------------------------------------------------------------------------------
PV1047    yoghurt          4745    6.8672995        1262.58          0    720.0000000
GV1047    yoghurt          4745    0.1288754   187.7863750          0    165.0000000
FV1047    yoghurt          4745   58.2347299        8213.61          0        3245.54
PQ1047    yoghurt          4745    0.7881413   143.9653867          0     60.8108108
GQ1047    yoghurt          4745    0.0148868    20.9010465          0     20.0000000
FQ1047    yoghurt          4745    9.2496966        1927.15          0        1500.00
VEXP1047  yoghurt          4745   65.2309048        8208.90          0        3245.54
VQ1047    yoghurt          4745   10.0527248        1925.39          0        1500.00
PV1050    oils             4745  109.0322872        4464.62          0        1250.00
GV1050    oils             4745    0.0509114   123.8891262          0    250.0000000
VEXP1050  oils             4745  109.0831986        4469.63          0        1250.00
PQ1050    oils             4745    0.2026191     8.1664335          0      2.0833333
GQ1050    oils             4745  0.000095033     0.2399075          0      0.5000000
VQ1050    oils             4745    0.2027142     8.1777837          0      2.0833333
VEXP1060  meats            4745  145.2190632       12160.28          0        4400.00
PV1061    mutton lamb goat 4745   46.7001514        5875.66          0        2000.00
GV1061    mutton lamb goat 4745    4.0667355        2898.32          0        2333.33
FV1061    mutton lamb goat 4745    1.7790082        1491.88          0        1125.00
PQ1061    mutton lamb goat 4745    0.9766703   118.4891463          0     40.0000000
GQ1061    mutton lamb goat 4745    0.0949154    71.9427262          0     58.3333333
FQ1061    mutton lamb goat 4745    0.0405620    33.8473991          0     22.5000000
VEXP1061  mutton lamb goat 4745   52.5458952        6736.05          0        2383.33
VQ1061    mutton lamb goat 4745    1.1121477   142.8849340          0     59.5833333
PV1062    beef buffaloe    4745   71.8885673        7786.15          0        4400.00
GV1062    beef buffaloe    4745    1.4674612        1516.48          0        1866.67
FV1062    beef buffaloe    4745    0.2887338   285.0930794          0    330.0000000
PQ1062    beef buffaloe    4745    2.7341218   293.2842596          0    162.9629630
GQ1062    beef buffaloe    4745    0.0593264    59.8049976          0     66.6666667
FQ1062    beef buffaloe    4745    0.0116464    11.2549787          0     15.0000000
VEXP1062  beef buffaloe    4745   73.6447624        8015.15          0        4400.00
VQ1062    beef buffaloe    4745    2.8050946   302.7680671          0    162.9629630
PV1063    fish             4745   17.0364713        3640.94          0        2000.00
GV1063    fish             4745    0.2207932   615.1999415          0        1000.00
FV1063    fish             4745    1.7711411        1766.42          0        1276.45
PQ1063    fish             4745    0.5013150   110.1752485          0     60.0000000
GQ1063    fish             4745    0.0058105    15.5568400          0     25.0000000
FQ1063    fish             4745    0.0689071    77.6724421          0     75.0000000
VEXP1063  fish             4745   19.0284056        4138.53          0        2000.00
VQ1063    fish             4745    0.5760326   137.3165226          0     75.0000000
VEXP1070  poultry and eggs 4745   56.9942965        5330.75          0        1500.00
PV1071    chicken          4745   19.0869208        2272.72          0    800.0000000
GV1071    chicken          4745    0.0442358    64.9881779          0     80.0000000
FV1071    chicken          4745   11.4196195        3073.02          0        1500.00
PQ1071    chicken          4745    0.5196332    62.4292985          0     21.9354839
GQ1071    chicken          4745    0.0014374     2.6079042          0      3.3333333
FQ1071    chicken          4745    0.3894337   256.4648253          0    300.0000000
VEXP1071  chicken          4745   30.5507761        3807.97          0        1500.00
VQ1071    chicken          4745    0.9105042   263.0449178          0    300.0000000
PV1072    egg              4745   10.6501241        1305.64          0    960.0000000
GV1072    egg              4745    0.0274176    37.0354228          0     50.0000000
FV1072    egg              4745   15.7659788        2877.51          0        1000.00
PQ1072    egg              4745    0.0597201     7.1453169          0      5.3333333
GQ1072    egg              4745  0.000142205     0.1957282          0      0.3472222
FQ1072    egg              4745    1.5101590   641.1348820          0    440.0000000
VEXP1072  egg              4745   26.4435205        3063.77          0        1000.00
VQ1072    egg              4745    1.5700212   640.8691381          0    440.0000000
VEXP1080  fruit            4745   61.9808870        5080.85          0        2342.50
PV1081    bananas          4745    9.6238989   981.5843573          0    410.0000000
GV1081    bananas          4745    0.0033519     3.0452340          0      6.7956349
FV1081    bananas          4745    0.0813285   140.5836966          0    540.0000000
PQ1081    bananas          4745   12.8368488        1335.24          0    535.8851675
GQ1081    bananas          4745    0.0036820     3.6073812          0      8.3333333
FQ1081    bananas          4745    0.1428339   266.0958323          0        1080.00
VEXP1081  bananas          4745    9.7085793   990.3909748          0    540.0000000
VQ1081    bananas          4745   12.9833647        1359.61          0        1080.00
PV1082    citrus fruits    4745   10.8132807   850.7694238          0    400.0000000
GV1082    citrus fruits    4745    0.0182116    35.8795585          0     63.2801299
FV1082    citrus fruits    4745    1.1410936        1128.14          0        2000.00
```

69

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
PQ1082	citrus fruits	4745	15.2475841	1190.99	0	538.3177570
GQ1082	citrus fruits	4745	0.0210877	31.9842346	0	83.3333333
FQ1082	citrus fruits	4745	7.3129317	20162.46	0	36000.00
VEXP1082	citrus fruits	4745	11.9725859	1393.91	0	2008.33
VQ1082	citrus fruits	4745	22.5816035	20183.09	0	36000.00
PV1083	mango	4745	12.9370599	1004.19	0	500.0000000
GV1083	mango	4745	0.1781787	251.6700496	0	384.2592593
FV1083	mango	4745	3.2593369	1804.03	0	1250.00
PQ1083	mango	4745	1.1335106	86.5742462	0	23.1445087
GQ1083	mango	4745	0.0163943	26.0661719	0	41.6666667
FQ1083	mango	4745	0.5257811	394.6165723	0	300.0000000
VEXP1083	mango	4745	16.3745754	2034.53	0	1250.00
VQ1083	mango	4745	1.6756859	401.6257403	0	300.0000000
PV1085	melon	4745	6.1235062	563.9875559	0	250.0000000
GV1085	melon	4745	0.0267360	21.6654095	0	16.6666667
FV1085	melon	4745	0.8244705	400.3162488	0	200.0000000
PQ1085	melon	4745	1.5011803	146.5806081	0	116.6666667
GQ1085	melon	4745	0.0068688	5.3175260	0	5.0000000
FQ1085	melon	4745	0.2960163	129.1821375	0	50.0000000
VEXP1085	melon	4745	6.9747127	684.2976597	0	250.0000000
VQ1085	melon	4745	1.8040653	193.0082464	0	116.6666667
PV1180	canned food	4745	0.5479885	460.2975559	0	500.0000000
GV1180	canned food	4745	0.0010988	9.3302735	0	30.8333333
VEXP1180	canned food	4745	0.5490873	460.4042949	0	500.0000000
PQ1180	canned food	4745	0.0483374	44.0558920	0	52.0673813
GQ1180	canned food	4745	0.000068948	0.5071063	0	1.6666667
VQ1180	canned food	4745	0.0484063	44.0610229	0	52.0673813
PV1089	other	4745	14.4112096	1678.11	0	1000.00
GV1089	other	4745	0.0452832	33.2276706	0	33.3333333
FV1089	other	4745	2.4939410	1506.41	0	1800.00
PQ1089	other	4745	1.4991386	297.3909975	0	272.7272727
GQ1089	other	4745	0.0047032	3.2296971	0	4.1666667
FQ1089	other	4745	0.2917836	160.2333889	0	150.0000000
VEXP1089	other	4745	16.9504337	2275.33	0	1900.00
VQ1089	other	4745	1.7956254	351.6371042	0	299.7272727
PV1090	vegetables	4745	139.7617482	6749.47	0	3000.00
GV1090	vegetables	4745	0.1223049	181.5563013	0	320.0000048
FV1090	vegetables	4745	9.4347901	2433.30	0	1200.00
PQ1090	vegetables	4745	25.4044420	1373.16	0	705.8823529
GQ1090	vegetables	4745	0.0197074	29.3325176	0	50.0000000
FQ1090	vegetables	4745	6.2294512	15111.35	0	20020.00
VEXP1090	vegetables	4745	149.3188432	7030.44	0	3000.00
VQ1090	vegetables	4745	31.6536006	15150.76	0	20020.00
PV1100	spices (and condiments)	4745	24.1752649	1881.36	0	1000.00
GV1100	spices (and condiments)	4745	0.2444157	566.9508989	0	631.9444443
FV1100	spices (and condiments)	4745	1.5022365	703.6396775	0	360.0000000
PQ1100	spices (and condiments)	4745	11.2171574	1017.13	0	350.3649635
GQ1100	spices (and condiments)	4745	0.0919247	196.9768919	0	166.6666667
FQ1100	spices (and condiments)	4745	2.7902916	6129.02	0	10000.00
VEXP1100	spices (and condiments)	4745	25.9219171	2061.49	0	1000.00
VQ1100	spices (and condiments)	4745	14.0993737	6218.53	0	10007.73
VEXP1110	sugar	4745	111.5378981	5533.45	0	2300.00
PV1111	refined sugar	4745	92.3029600	4635.28	0	1590.00
GV1111	refined sugar	4745	0.0261195	39.1551207	0	55.0000000
VEXP1111	refined sugar	4745	92.3290795	4635.80	0	1590.00
PQ1111	refined sugar	4745	7.9706355	399.3761739	0	144.5454545
GQ1111	refined sugar	4745	0.0023213	3.5158378	0	5.0000000
VQ1111	refined sugar	4745	7.9729568	399.4335643	0	144.5454545
PV1112	desi sugar (gur)	4745	14.3993396	2137.50	0	800.0000000
GV1112	desi sugar (gur)	4745	0.0310418	23.5070417	0	11.6666667
FV1112	desi sugar (gur)	4745	4.7784373	2284.17	0	1200.00
PQ1112	desi sugar (gur)	4745	1.8425442	250.0131820	0	114.2857143
GQ1112	desi sugar (gur)	4745	0.0042795	3.3258041	0	1.6666667
FQ1112	desi sugar (gur)	4745	1.7637415	1909.49	0	1500.00
VEXP1112	desi sugar (gur)	4745	19.2088186	3074.70	0	1200.00
VQ1112	desi sugar (gur)	4745	3.6105652	1922.24	0	1503.00
VEXP1120	tea and coffee	4745	55.0318180	3446.22	0	2000.00
PV1121	tea	4745	54.4359467	3358.17	0	2000.00

```
Variable   Label                                          N         Mean      Std Dev    Minimum      Maximum
--------------------------------------------------------------------------------------------------------------
GV1121     tea                                          4745    0.1871055  646.4742240         0   822.9166667
VEXP1121   tea                                          4745   54.6230522     3414.75          0      2000.00
PQ1121     tea                                          4745    8.9004669  539.0429782         0   296.2962963
GQ1121     tea                                          4745    0.0377934  130.9270788         0   166.6666667
VQ1121     tea                                          4745    8.9382603  553.8189871         0   296.2962963
PV1122     coffee                                       4745    0.4032609  323.9178284         0   240.0000000
GV1122     coffee                                       4745    0.0055049   12.4663992         0    11.4583333
VEXP1122   coffee                                       4745    0.4087658  324.6945025         0   240.0000000
PQ1122     coffee                                       4745    0.0073279    5.8622349         0     3.0303030
GQ1122     coffee                                       4745  0.000160002    0.4080576         0     0.4166667
VQ1122     coffee                                       4745    0.0074879    5.8869464         0     3.0303030
PV1130     bottled drinks (cola squash etc)             4745   13.5012643     1289.58          0   800.0000000
GV1130     bottled drinks (cola squash etc)             4745    0.0057744   10.6457275         0    14.2424242
VEXP1130   bottled drinks (cola squash etc)             4745   13.5070387     1289.82          0   800.0000000
PQ1130     bottled drinks (cola squash etc)             4745    1.2979191  131.2643715         0    60.0000000
GQ1130     bottled drinks (cola squash etc)             4745  0.000256156    0.3933880         0     0.4166667
VQ1130     bottled drinks (cola squash etc)             4745    1.2981753  131.2638128         0    60.0000000
PV1150     tobacco cigarettes naswar pan                4745   70.6434651     5493.73          0      1500.00
GV1150     tobacco cigarettes naswar pan                4745    0.1046792   95.5979943         0    60.0000000
VEXP1150   tobacco cigarettes naswar pan                4745   70.7481443     5491.62          0      1500.00
VEXP1190   other foods                                  4745   38.6716719    34077.09          0     58652.63
PV1191     ground nuts                                  4745    4.3612098  597.3008097         0   500.0000000
GV1191     ground nuts                                  4745    0.0269082   42.4594578         0    29.5151515
FV1191     ground nuts                                  4745    1.1970490  788.3124190         0   300.0000000
PQ1191     ground nuts                                  4745    0.2732522   39.4098197         0    50.0000000
GQ1191     ground nuts                                  4745    0.0015131    2.3852312         0     1.6666667
FQ1191     ground nuts                                  4745    0.1664363  193.4405631         0   150.0000000
VEXP1191   ground nuts                                  4745    5.5851670  981.1320580         0   500.0000000
VQ1191     ground nuts                                  4745    0.4412016  196.8844880         0   150.0000000
PV1199     miscellaneous food expenses                  4745   24.1755403     3273.64          0      1500.00
GV1199     miscellaneous food expenses                  4745            0            0         0            0
FV1199     miscellaneous food expenses                  4745    8.9109645    33835.83          0     58552.63
VEXP1199   miscellaneous food expenses                  4745   33.0865048    34034.03          0     58652.63
VEXP2000   fuel and lighting                            4745  157.8015569     9433.23          0      3520.00
VEXP2001   expenditure on energy                        4745  116.5263906     8788.68          0      3500.00
PV2113     kerosene matches and candles                 4745   41.2634259     3264.83          0      1500.00
GV2113     kerosene matches and candles                 4745    0.0117403   12.5845213         0     6.0000000
VEXP2113   kerosene matches and candles                 4745   41.2751663     3264.93          0      1500.00
VEXP3000   personal use items                           4745  304.6889941    21282.41          0     17055.00
VEXP3210   clothing                                     4745  172.1202837    11577.31          0      4166.67
PV3211     children clothing and material               4745   59.8524160     4939.50          0      1666.67
GV3211     children clothing and material               4745    2.7417410  680.4856684         0   500.0000000
VEXP3211   children clothing and material               4745   62.5941570     5045.19          0      1666.67
PV3212     adult clothing and material                  4745  105.8841968     8259.17          0      4166.67
GV3212     adult clothing and material                  4745    3.6419299     1093.83          0   625.0000000
VEXP3212   adult clothing and material                  4745  109.5261267     8376.07          0      4166.67
VEXP3220   Footwear                                     4745   63.9753008     4223.85          0      2500.00
PV3221     children footwear                            4745   22.5628283     2007.32          0      1000.00
GV3221     children footwear                            4745    0.2176260  142.5978541         0    83.3333333
VEXP3221   children footwear                            4745   22.7804543     2015.43          0      1000.00
PV3222     adult footwear                               4745   40.8385858     3146.36          0      2500.00
GV3222     adult footwear                               4745    0.3562607  213.7240526         0   250.0000000
VEXP3222   adult footwear                               4745   41.1948464     3159.22          0      2500.00
PV3230     other personal effects                       4745   11.9099614     1990.39          0      1250.00
GV3230     other personal effects                       4745    0.1389327  185.9727197         0   250.0000000
VEXP3230   other personal effects                       4745   12.0488941     2003.08          0      1250.00
PV3240     stitching or repair of wearing apparel       4745   26.4936133     2262.61          0      1666.67
GV3240     stitching or repair of wearing apparel       4745    0.2249068  181.5728388         0   166.6666667
VEXP3240   stitching or repair of wearing apparel       4745   26.7185201     2265.58          0      1666.67
PV3320     household textiles                           4745   12.3097638     1739.47          0      1000.00
GV3320     household textiles                           4745    0.1269520  147.5634110         0   166.6666667
VEXP3320   household textiles                           4745   12.4367157     1748.65          0      1000.00
VEXP4000   housing                                      4745  564.5786571    79423.37          0     40291.67
VEXP4210   rent and housing expenitures                 4745  420.2684332    48163.60          0     40291.67
VEXP4211   actual rent                                  4745   24.9206842     9497.13          0     12000.00
VEXP4213   imputed rent                                 4745  327.7078014    43191.18          0     40000.00
PV4214     repair and maintenance of house              4745   44.4257914    15073.39          0      8333.33
GV4214     repair and maintenance of house              4745    0.8449605     2026.20          0      2083.33
```

71

```
Variable    Label                                    N       Mean        Std Dev      Minimum     Maximum
------------------------------------------------------------------------------------------------------------
VEXP4214    repair and maintenance of house          4745    45.2707519   15282.45           0    8333.33
PV4215      housing and property taxes               4745     2.0810292    1530.45           0    5416.67
GV4215      housing and property taxes               4745     0.0022636    9.5130310         0    14.5833333
VEXP4215    housing and property taxes               4745     2.0832928    1530.53           0    5416.67
VEXP4216    annual garbage disposal expenditure      4745     1.9641110    658.6763173       0    500.0000000
VEXP4217    annual water expenditure                 4745     5.9345026    1419.16           0    2816.67
VEXP4219    annual utility repairs                   4745    12.3872893    4189.22           0    4166.67
PV4230      repair and servicing of hh effects       4745     3.9177642    1748.87           0    6666.67
GV4230      repair and servicing of hh effects       4745     0.000925116  3.2506244         0    4.1666667
VEXP4230    repair and servicing of hh effects       4745     3.9186893    1748.87           0    6666.67
PV4240      other household effects                  4745     1.8968885    474.4250035       0    750.0000000
GV4240      other household effects                  4745     0.0023923    10.6793374        0    50.0000000
VEXP4240    other household effects                  4745     1.8992808    474.5640022       0    750.0000000
PV4320      kitchen equipment incl crockery          4745    11.8018308    1695.95           0    1333.33
GV4320      kitchen equipment incl crockery          4745     0.1777230    401.4772700       0    708.3333333
VEXP4320    kitchen equipment incl crockery          4745    11.9795537    1747.42           0    1333.33
PV4330      furniture and fittings                   4745     7.7057945    4560.15           0    5833.33
GV4330      furniture and fittings                   4745     0.2273297    466.7159824       0    833.3333333
VEXP4330    furniture and fittings                   4745     7.9331242    4602.66           0    5833.33
VEXP4390    other durable housing expenditure        4745   118.5795758    58380.75          0    37500.00
PV4398      home improvements and additions          4745    56.5486984    32367.36          0    21666.67
GV4398      home improvements and additions          4745     0.0365786    84.6372383        0    83.3333333
VEXP4398    home improvements and additions          4745    56.5852770    32367.84          0    21666.67
PV4399      land/buildings for residence/investment  4745    61.9942987    46876.67          0    37500.00
GV4399      land/buildings for residence/investment  4745            0             0         0           0
VEXP4399    land/buildings for residence/investment  4745    61.9942987    46876.67          0    37500.00
VEXP1160    hh food/clothing recd from emper         4745    50.9999502    10836.66          0    9225.00
VEXP5000    miscellaneous                            4745     1269.72      126297.99   4.1666667  56422.67
VEXP5110    toiletries                               4745   109.1913881    5845.43           0    2600.00
PV5111      commercial or handmade soap              4745    54.5244182    2921.20           0    1300.00
GV5111      commercial or handmade soap              4745     0.0712759    115.5320773       0    100.0000000
VEXP5111    commercial or handmade soap              4745    54.5956941    2922.72           0    1300.00
PV5119      oth pers care (cosmtcs  soap cmbs etc)   4745    54.5244182    2921.20           0    1300.00
GV5119      oth pers care (cosmtcs  soap cmbs etc)   4745     0.0712759    115.5320773       0    100.0000000
VEXP5119    oth pers care (cosmtcs  soap cmbs etc)   4745    54.5956941    2922.72           0    1300.00
PV5120      personal services (eg haircut shoeshine) 4745    15.0560800    1297.25           0    1250.00
GV5120      personal services (eg haircut shoeshine) 4745     0.0371588    54.9208363        0    50.0000000
VEXP5120    personal services (eg haircut shoeshine) 4745    15.0932387    1300.73           0    1250.00
VEXP5130    recreation and travel                    4745    67.4051826    12526.68          0    7961.67
PV5131      newspapers books and other entertainment 4745    10.0780347    2205.20           0    1000.00
GV5131      newspapers books and other entertainment 4745     0.0497442    148.1750718       0    400.0000000
VEXP5131    newspapers books and other entertainment 4745    10.1277789    2211.39           0    1000.00
PV5133      recreation personal travel lodging       4745    15.8107162    7986.26           0    7916.67
GV5133      recreation personal travel lodging       4745     0.1410642    178.6775176       0    416.6666667
VEXP5133    recreation personal travel lodging       4745    15.9517804    7991.45           0    7916.67
PV5134      meals eaten outside the house            4745    39.7120400    8357.47           0    3000.00
GV5134      meals eaten outside the house            4745     1.6135834    1443.41           0    800.0000000
VEXP5134    meals eaten outside the house            4745    41.3256233    8494.01           0    3000.00
VEXP5140    personal transport expenses              4745    72.5146283    14532.75          0    6250.00
PV5141      gas motor oil for personal transport     4745    25.4146283    9455.43           0    5000.00
GV5141      gas motor oil for personal transport     4745     0.1701788    622.0581936       0    1000.00
VEXP5141    gas motor oil for personal transport     4745    25.5848072    9475.27           0    5000.00
PV5142      repair/service of vehicles excl gas+oil  4745     9.3157996    6007.85           0    3333.33
GV5142      repair/service of vehicles excl gas+oil  4745     0.0064835    76.9772589        0    333.3333333
VEXP5142    repair/service of vehicles excl gas+oil  4745     9.3222831    6014.72           0    3333.33
PV5144      public transport incl rickshaws+taxis    4745    31.8313216    2925.69           0    1250.00
GV5144      public transport incl rickshaws+taxis    4745     0.0879436    88.7173199        0    50.0000000
VEXP5144    public transport incl rickshaws+taxis    4745    31.9192652    2929.14           0    1250.00
VEXP5190    misc frequently incurred expenditure     4745    33.2837763    12569.98          0    26400.00
PV5191      wages to servants gardeners etc          4745    14.8354800    6949.95           0    9800.00
GV5191      wages to servants gardeners etc          4745     0.0012142    3.5714174         0    4.0000000
VEXP5191    wages to servants gardeners etc          4745    14.8366942    6949.95           0    9800.00
PV5192      postal articles telegram telephone       4745     8.6464118    4108.87           0    3232.33
GV5192      postal articles telegram telephone       4745     0.0173427    33.9199978        0    41.6666667
VEXP5192    postal articles telegram telephone       4745     8.6637545    4109.07           0    3232.33
VEXP5193    annual telephone expenditure             4745     9.7833277    5772.04           0    25000.00
VEXP5210    health expenses                          4745   453.3843135    74712.68          0    55506.00
VEXP5213    non-diah. health services                4745   419.0125078    73275.08          0    55200.00
```

```
Variable  Label                                         N        Mean      Std Dev      Minimum      Maximum
-------------------------------------------------------------------------------------------------------------
VEXP5217  diah. health services                       4745   34.3718058    10096.04           0      7005.00
VEXP5240  education                                   4745  162.2584857    24232.71           0     33333.33
VEXP5249  received by household in scholarships       4745    1.1206481     1194.68           0      1666.67
VEXP5250  help received for educational expenses      4745    2.1675036     1937.55           0      2250.00
VEXP5241  adminn/regn/tuition                         4745   19.3612064     3867.71           0      4250.00
VEXP5242  uniforms                                    4745   21.7603504     2034.36           0   583.3333333
VEXP5243  books for education                         4745   18.8313393     1995.64           0   750.0000000
VEXP5244  transport for education                     4745    5.6712761     1801.22           0   833.3333333
VEXP5245  private tuition                             4745   11.7965680     3071.30           0      3500.00
VEXP5246  exam fees                                   4745    2.5334060   502.5561765          0   275.0000000
VEXP5247  other education expenditure                 4745   18.2405140     3787.91           0      4600.83
VEXP5248  unspecified education expenditure           4745   27.4662868     9486.79           0     16666.67
PV5251    ed/profnl services repted in consn sctn     4745   32.8650701     9686.88           0     16666.67
GV5251    ed/profnl services repted in consn sctn     4745    0.4443169     1502.45           0      2500.00
VEXP5251  ed/profnl services repted in consn sctn     4745   33.3093870    10241.77           0     16666.67
PV5260    stationery books (non-education-related)     4745    3.9929373     1120.72           0   583.3333333
GV5260    stationery books (non-education-related)     4745    0.0120749    59.4775016          0   166.6666667
VEXP5260  stationery books (non-education-related)     4745    4.0050122     1122.26           0   583.3333333
VEXP5290  misc infrequently incurred expenditure      4745  327.4014430    65283.72           0     25666.67
PV5291    cash losses                                 4745   31.5811828    20756.91           0     13333.33
GV5291    cash losses                                 4745            0            0           0            0
VEXP5291  cash losses                                 4745   31.5811828    20756.91           0     13333.33
PV5292    marriages births and other ceremonies       4745  144.1199721    32452.83           0     12500.00
GV5292    marriages births and other ceremonies       4745    4.8640622     3900.88           0      2500.00
VEXP5292  marriages births and other ceremonies       4745  148.9840344    33685.50           0     12500.00
PV5293    funerals and related death expenses         4745   24.6153422     5148.68           0      1666.67
GV5293    funerals and related death expenses         4745    1.1073991     1578.77           0      1666.67
VEXP5293  funerals and related death expenses         4745   25.7227413     5414.16           0      1666.67
PV5294    legal expenses                              4745   14.2272278     9448.51           0      8166.67
GV5294    legal expenses                              4745            0            0           0            0
VEXP5294  legal expenses                              4745   14.2272278     9448.51           0      8166.67
VEXP5295  remittances to household members            4745    5.4119619     7180.02           0     13333.33
PV5299    dowry                                       4745   93.6649455    40258.78           0     25000.00
GV5299    dowry                                       4745    7.8093493    12584.98           0     12500.00
VEXP5299  dowry                                       4745  101.4742948    42375.98           0     25000.00
VEXP5300  miscellaneous durable epxenses              4745   27.0009083    18455.21           0     25000.00
PV5321    radio                                       4745    0.4800579   370.7239430          0   183.3333333
GV5321    radio                                       4745    0.0813695   115.7043821          0   100.0000000
VEXP5321  radio                                       4745    0.5614274   388.0844502          0   183.3333333
PV5326    gramophone/phonograph/tape-player           4745    1.6728288     1519.83           0      2083.33
GV5326    gramophone/phonograph/tape-player           4745    0.1965904   270.2528704          0   208.3333333
VEXP5326  gramophone/phonograph/tape-player           4745    1.8694192     1543.09           0      2083.33
PV5331    camera                                      4745    0.2047255   238.2909125          0   166.6666667
GV5331    camera                                      4745    0.0972397   210.7375646          0   166.6666667
VEXP5331  camera                                      4745    0.3019652   317.9366499          0   166.6666667
PV3312    jewelry                                     4745   12.6844331     7795.87           0      3600.00
GV3312    jewelry                                     4745    4.7048466     6777.67           0     16666.67
VEXP3312  jewelry                                     4745   17.3892797    10315.76           0     16666.67
PV5332    guns                                        4745    0.4752515   649.9265931          0   500.0000000
GV5332    guns                                        4745    0.0056173    32.0422777          0    66.6666667
VEXP5332  guns                                        4745    0.4808689   650.7047290          0   500.0000000
PV5341    bicycle                                     4745    2.0225462   864.5585996          0   500.0000000
GV5341    bicycle                                     4745    0.0371667    85.5924686          0    83.3333333
VEXP5341  bicycle                                     4745    2.0597129   873.8172646          0   500.0000000
PV5342    motorcycle/scooter                          4745    9.4537009     7140.02           0      3000.00
GV5342    motorcycle/scooter                          4745            0            0           0            0
VEXP5342  motorcycle/scooter                          4745    9.4537009     7140.02           0      3000.00
PV5349    motor rickshaw                              4745            0            0           0            0
GV5349    motor rickshaw                              4745            0            0           0            0
VEXP5349  motor rickshaw                              4745            0            0           0            0
PV5343    automobile/truck                            4745    7.3853679    13853.74           0     25000.00
GV5343    automobile/truck                            4745            0            0           0            0
VEXP5343  automobile/truck                            4745    7.3853679    13853.74           0     25000.00
PV5399    other durables                              4745    4.4375795     9316.35           0     11416.67
GV5399    other durables                              4745    0.4508663   714.2151254          0   679.1666667
VEXP5399  other durables                              4745    4.8884458     9349.12           0     11416.67
VEXP5149  transport subsidy from employer             4745    5.7105050     4600.12           0      4166.67
V2XP4213  Regressed imputed rent                      4745  500.1080588    31235.79           0      4450.01
```

73

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
V2XP1600	Housing benefits in kind S5Bq20	4745	21.1322469	11069.37	0	10000.00
V3XP1600	Transport benefits in kind S5Bq22	4745	5.7105050	4600.12	0	4166.67
V4XP1600	Other benefits in kind S5Bq24	4745	1.9539750	1382.90	0	1416.67
V2XP5321	y flow: radio	4745	0.9834616	223.7694067	0	146.9681981
V2XP5326	y flow: gramphne/phongrph/tp-plyr	4745	4.2021764	1006.32	0	2250.00
V2XP5331	y flow: camera	4745	0.5218665	203.8051163	0	123.9407991
V2XP5332	y flow: guns	4745	2.9779186	1554.68	0	1022.59
V2XP5341	y flow: bicycle	4745	4.1354257	1827.30	0	3299.26
V2XP5342	y flow: motorcycle/scooter	4745	7.0353685	4756.10	0	5390.61
V2XP5349	y flow: motor rickshaw	4745	0.6459429	901.5212117	0	1005.22
V2XP5343	y flow: automobile/truck	4745	8.1129114	8189.78	0	11183.03
V2XP5399	y flow: other durables	4745	8.5159118	4868.54	0	4304.81
V2XP5300	Income flow from misc durables	4745	27.0009083	18455.21	0	25000.00
V2XP5295	All remits (incl. to non-hh membrs)	4745	27.6674775	10650.69	0	13333.33
GOOD	1 if passes cleaning test	4745	1.0000000	0	1.0000000	1.0000000

The program below is the main progr am in the first stage of the construction of PIHSEXPN. It contains comments which are helpful in understanding the creation of the expenditure variables. This file is included in the PIHSEXPN.WP 5 document as well as the macros which are referred to in this program.

PCREATE SAS PROGRAM:

```
/*
%macro ob;
obs=10
%mend;
*/
%macprint;

%macro cm;
proc contents;
proc means;
%mend;
```

/*the ob macro enables a quick run through for de-bugging. the macprint
macro is equivalent to typing 'option mprint' and will give a log showing
the programme generated by the macros*/

/*THIS PROGRAM CREATES PIHSEXP, A SAS DATA SET WITH THE EXPENDITURE
INFORMATION FROM THE PIHS. IT WILL EVENTUALLY USE ONLY RAW DATA SETS
PLUS CONSUTRUCTED PSU LEVEL PRICE DATA SETS. IT USES A NUMBER OF MACROS
AS INDICATED. PIHSEXP WILL BE FED INTO THE PROGRAMME RENAME TO CREATE
PIHSEXPN, A SAS DATA SET WITH VARIABLE NUMBERING ANALOGOUS TO THE HIES
DATA SETS*/

/*FIRST WE GET HHSIZE FROM A WORLD BANK PROGRAM ROSTER1A.SAS. THIS OUTPUTS
HHSIZE AND HHSIZE WEIGHTED BY RESIDENCE (hhsizew). WE ALSO ADD TO THIS PROGRAMME
TO GET VARIABLES HHSIZE2 AND HHSIZE3 WHICH ARE DIFFERENT DEFINITIONS OF THE
HOUSEHOLD(HHSIZE2=HHSIZE APPROXIMATING THE HIES DEFN (ABSENCE LESS THAN SIX
MONTHS UNLESS PERSON IS HEAD) AND HHSIZE3 WHICH ONLY INCLUDES MEMBERS WITH NO
ABSENCE) THE PROGRAM USES F01A.SSP ASSUMED TO BE ON C:\O\PAK\DATA*/

*%HHSIZE;

/*NOW ADD SAMPLING WEIGHTS AS WELL AS PROVINCE AND URBAN/RURAL INDICATORS.
THE WEIGHTS ARE CONTAINED IN THE SAS EXPORT DATA SET WEIGHTS.EXP
ASSUMED TO BE ON C:\O\PAK\DATA. THE PROVINCE AND URBAN/RURAL INDICATORS ARE
ALSO IN THIS DATA SET THOUGH THEY COULD BE JUST AS WELL CALCULATED FROM
HHCODE (FIRST TWO DIGITS). WEIGHTS.EXP IS AT THE CLUSTER LEVEL. IT IS CONVERTED
TO HHOLD LEVEL USING F01A.SSP. IF A HHOLD IS IN THE FINAL DATA SET IT MUST
BE IN THIS ONE*/

*%WEIGHTS;

/*ADD ASCII DATA SET on ENERGY. THIS IS NAMED ENERGY.OUT AND IS ASSUMED

TO BE ON C:\O\PAK\DATA. THIS WAS CALCULATED WITHIN THE BANK USING
PROGRAMMES SINCE LOST.*/

*%ENERGY;

/* ADD ACTUAL, IMPUTED (BOTH REPORTED) AND REGRESSED RENTS
FROM RAW DATA SETS F02A, F02B AND F02C, ASSUMED TO BE ON C:\O\PAK\DATA */

*%Hrent;

/*THIS PART OF THE PROGRAM PUTS ALL THE SECTION 11 GOODS IN TWO
SEPARATE DATA SETS (A AND B), DISTINGUISHING BETWEEN PURCHASED AND
GIFTED. IT DOES NO CLEANING - SO IT PRODUCES IDENTICAL RESULTS TO THE
DATA INPUT INTO AGGRNK
IT USES F11A.SSP AND F11B.SSP, BOTH SAS XPORT DATA SETS ASSUMED TO
BE ON C:\O\PAK\DATA*/

*%F11AB;

/*THIS MACRO PRODUCES IN-KIND PAYMENTS FROM PRIMARY EMPLOYMENT. IT IS
SUBSTANTIVELY IDENTICAL TO THE WB PROGRAM. IT USES DATA (F05B2 AND F05B3)
ASSUMED TO BE ON C:\O\PAK\DATA TO PRODUCE PG.INKIND*/

*%INKIND;

/*THIS MACRO CALCULATES THE EXPENDITURE WITHIN THE MOST RECENT YEAR ON
MISCELLANEOUS DURABLES. YOU NEED F11C.SSP TO RUN IT. THIS IS ASSUMED TO BE
LOCATED ON C:\O\PAK\DATA*/

*%PURDUR;

/*THIS MACRO IS A BANK ONE WHICH CALCULATES THE DEPRECIATION IN EACH DURABLE
TO GIVE A FLOW MEASURE OF VALUE OF DURABLES (EXCL JEWELLERY). IT USES
F11C.SSP TO RUN IT. THIS IS ASSUMED TO BE LOCATED ON C:\O\PAK\DATA*/

*%DEPDUR;

/*THIS MACRO GIVES EDN SUB-TOTALS.
DATA (F03B2) ASSUMED TO BE ON C:\O\PAK\DATA
CLEANING RULES SAME AS WORLD BANK PROGRAM EXED3B2B.SAS
NOTE NEED TO DO SOMETHING ABOUT SUB-COMPONENTS
*/

*%EDUCATE;

/*THIS MACRO GIVES TOTAL REMITTANCES, TO HOUSEHOLD MEMBERS AND TO NON
HHOLD MEMBERS DATA (F16A2) ASSUMED TO BE ON C:\O\PAK\DATA
FILE USES NO CLEANING AND SAME DECISION RULES AS WB PROGRAM
*/

*%REMIT;

/*THIS MACRO GIVES UTILITY EXPENDITURES.
DATA F02C ASSUMED TO BE ON C:\O\PAK\DATA*/

*%UTILS;

/*THIS MACRO GIVES THE QUANTITY OF HOME CONSUMPTION FOR ALL FOODSTUFFS
SELF-PRODUCED. IT USES CLEANING IDENTICAL TO EARLIER BANK PROGRAM:
IE IT PRODUCES IDENTICAL DATA SET TO THAT PRODUCED BY THE BANK
IT USES EXPORT FILE F12B2.SSP ASSUMED TO BE ON C:\O\PAK\DATA
*/

*%QHOME;

/*THIS PART OF THE PROGRAM PRODUCES VHOME, THE VALUE OF HOME
CONSUMPTION FOR ALL FOODSTUFFS SELF-PRODUCED. IT USES CLEANING
IDENTICAL TO THAT IN EARLIER BANK PROGRAM: IE IT PRODUCES AN
IDENTICAL DATA SET TO THAT PRODUCED BY THE BANK
IT USES XPORT FILE F12B2.SSP WHICH IS ASSUMED TO BE ON C:\O\PAK\DATA
*/

*%VHOME;

/*THIS PART OF THE PROGRAM ADDS VALUES (PURCHASED ONLY) FOR ALL
 FOODSTUFFS INCL MISC. THIS IS DONE USING F12A.SSP, ASSUMED TO BE ON THE
C:\O\PAK\DATA. ALSO USE DATA SET ED TO CALCULATE VALUES FOR FOODCODES
333, 334 AND 335 WHERE MISSING WHICH IS OFTEN DUE TO
QUESTIONNAIRE BEING MISTAKENLY BLACKED OUT. ED PROVIDES A VARIABLE ON
THE EDUCATION OF THE HEAD WHICH IS USED TO PREDICT EXPECTED VALUES OF
PURCHASES. THE LOCATION OF THE PROGRAM FOR CREATING ED IS UNKNOWN.
THE MACRO vpur PRODUCES AN INTERMEDIATE DATA SET PG.VPUR,
WHICH IS USED IN THE following STEP (QPUR)*/

*%VPUR;

/*NOW WE CREATE QUANTITIES PURCHASED USING PG.VPUR CREATED USING %VPUR ABOVE
AND PRICTOT.EXP, A SAS EXPORT DATA FILE WITH DATA ON PRICES ASSUMED TO BE
ON C:\O\PAK\DATA. THE QUANTITIES AND VALUES ARE COMBINED INTO QPUR.
NOTE WE DO NOT ATTEMPT TO GENERATE THESE PRICES FROM THE RAW DATA. SEE PAPER
FOR A DESCRIPTION OF THEM*/

*%QPUR;

/*THIS MACRO ADDS GIFTED (IN KIND) QUANTITIES FROM THE RAW DATA USING A WORLD
BANK PROGRAM IE substantively IDENTICAL DATA SET TO THAT PRODUCED BY THE BANK
DATA SET PRODUCED CALLED QINKIND
THE DATA (F12A.SSP) IS ASSUMED TO BE ON C:\O\PAK\DATA
NOTE QINKIND IS AN INTERMEDIATE DATA STEP USED TO GENERATE VINKIND - SEE BELOW
*/

```
*%QINKIND;

/*THIS MACRO TAKES IN-KIND QUANTITIES FROM QINKIND AND USES THE PRICES FROM
PRICTOT TO CONVERT THEM INTO VALUES. BOTH VALUES AND QUANTITIES ARE KEPT IN
VINKIND*/

*%VINKIND;

/*THIS MACRO GETS HEALTH EXPENDITURE INFO FROM THE HEALTH MODULE. THIS WAS DONE
BY LEAH GUTIEREZ AND WE USE HER OUTPUT NOT THE ORIGINAL DATA*/

%HEALTH;

/*THIS PART OF THE PROGRAM
COMBINES ALL THE NEW SUB DATA SETS WITH AGGRNK
ALTERNATIVELY (DEPENDING ON USE OF /* AND IT JUST UPDATES PIHSEXP BY
ANOTHER SUB DATA SET. NOTE PIHSEXP WILL BE USED TO CREATE PIHSEXPN WHICH
WILL CLEAN AND RELABEL AND CREATE SEVERAL NEW VARIABLES*/

%macro w;
where=(hhcode>0)
%mend;

DATA PG.PIHSEXP; MERGE
PG.HHSIZE  (%w)
PG.WEIGHTS (%w)
PG.TEMPENER(%w)
PG.HRENT(%w)
PG.S11A PG.S11B(%w)
PG.QHOME(%w)
PG.VHOME(%w)
PG.VINKIND(%w)
PG.QPUR(%w)
PG.DURPURCH(%w)
PG.DURDEP(%w)
PG.TEMPEDN(%w)
PG.REMIT(%w)
PG.TEMPUTIL(%w)
PG.INKIND(%w)
PG.HEALTH(%W);
BY HHCODE;
RUN;
/*
DATA PG.PIHSEXP; MERGE
PG.PIHSEXP
PG.HEALTH (%w);
BY HHCODE;
RUN;
*/
```