

Supplemental Information

Peru: Living Standards Measurement Survey 1994

Development Research Group
Poverty and Human Resources

The World Bank

August 1998

Table of Contents

1. Introduction	1
2. Survey Questionnaire	1
2.1 The Household Questionnaire	3
2.2 The Community Questionnaire	5
3. Sample	5
3.1 The Household Survey Sample	5
3.2 The Community Survey Sample	7
3.3 Expansion Factors	7
4. Survey Management and Fieldwork	8
5. Deflation of Monetary Values	8
6. Data Files	9
7. Subset of Panel Data -- 1991 and 1994 Data Sets	12
8. Data Quality	13
Appendix A: How to Obtain LSMS Survey Data and Documentation	15
Appendix B: Revised Expansion Factors for the 1991 Data	16
Appendix C: Price Indices to Compare Monetary Values in 1991 and 1994	17
Appendix D: Composition of Household Income and Expenditure Aggregates	18
Appendix E: Deflators used for 1994 ENNIV	20

List of Tables

Table 1: ENNIV 1994 Survey Documentation (Spanish)	2
Table 2: Sections of the 1994 Household Questionnaire	4
Table 3: Sections of the Community Questionnaire	5
Table 4: 1994 Sample -- Number of Segments and Dwellings, by Region	6
Table 5: 1994 LSMS - Expansion Factors, by Region and Sample Segment	7
Table 6: Peru 1994 LSMS - Household Questionnaire Data Files	10
Table 7: Peru 1994 LSMS - Additional Miscellaneous Files	11
Table 8: Peru 1994 LSMS -- Community Questionnaire Files	11
Table B.1: 1991 Expansion Factors by Region, Initial and Revised	16
Table E.1: Deflators used in ENNIV 1994 to Give Monetary Values to Prices in Each Domino and to 6/94	20

This document was prepared by Diane Steele (DECRG), Gillette Hall (consultant) and Donna MacIsaac (consultant) under the direction of Margaret Grosh (DECRG). The contribution of Cuánto is gratefully acknowledged.

1. Introduction

The *Living Standards Measurement Study (LSMS)* surveys were developed by the World Bank to provide policy relevant data on the welfare and living standards of households in developing countries. Four LSMS surveys have been conducted in Peru, where they are referred to as the *Encuesta Nacional de Hogares Sobre Medición de Vida (ENNIV)*. Geographic coverage of these surveys varies by year: the first survey, conducted in 1985-86, covers the entire country; the 1990 survey covers only Metropolitan Lima; the 1991 survey covers 70 percent of the country's population (omitting population groups living in areas that were considered too difficult to reach in that year due to terrorist activity); and, the 1994 survey again covers the entire country. The Peru LSMS survey data are available upon request from the World Bank. Procedures for obtaining data and additional documentation are described in Appendix A.

This document provides a general description of the 1994 LSMS for Peru. It is intended to be used in addition to the Basic Information Document (see Table 1). It briefly describes the survey sample, questionnaire, and fieldwork techniques, and provides a guide to the survey data. Drawing on the World Bank's experience to date, the document discusses some of the particular characteristics of this data set and suggests methods for handling them. The text is cross-referenced with the extensive set of Spanish-language documentation (listed in Table 1) which Spanish speakers can refer to for further guidance in using and interpreting the data. The document also contains information useful to researchers using both the 1991 and 1994 data.

The 1994 LSMS survey provides data for 3,623 households from a nationally-representative sample. The questionnaire is very similar to that used in prior survey years, and hence many results obtained are comparable over time. The Peru LSMS surveys contain a subset of panel data; panel data exist for Lima in all survey years, and for the entire survey in 1991 and 1994. This document provides some guidelines for identifying panel observations in the 1991 and 1994 data sets.

The 1994 LSMS in Peru was designed and conducted by Cuánto, S.A., a Peruvian research group, with technical and financial support from the World Bank and the Interamerican Development Bank.

2. Survey Questionnaire

The 1994 Peru LSMS includes both a household and community questionnaire. The household questionnaire collects socio-economic information about families and their individual members. Data collected cover diverse aspects of household welfare: health, education, consumption and income, etc. The community questionnaire, implemented only in rural areas, gathers information on population size, economic activity, and social infrastructure (schools, health centers, etc.). It also collects local market prices for basic consumption goods, in order to assess regional cost-of-living differentials. The community questionnaire was conducted in population centers where the household survey was also carried out, so the data from the two surveys can be linked by geographic area.

Table 1: ENNIV 1994 Survey Documentation (Spanish)

Title	Description
1. Información Básica	<i>Basic Information</i> -- Describes the basic design, coverage and characteristics of the survey.
2. Diseño Muestral	<i>Sample Design</i> -- Provides an in-depth description of the survey sample. Also includes original expansion factors, which have since been modified for greater accuracy (see Appendix B).
3. Cuestionario de la Encuesta de Hogares	<i>Household Survey Questionnaire</i>
4. Características y Equipamiento de Centros Poblados	<i>Community Questionnaire</i>
5. Especificaciones para la Composición del Ingreso y del Gasto	<i>Composition of Income and Expenditure Aggregates</i> -- Describes composition of household income and expenditure aggregates (this document may be also be used with reference to the 1991 data, for which aggregates were constructed using the same methodology).
6. Programa de la Encuesta (concepto y definiciones)	<i>Survey Plan (concepts and definitions)</i> -- Describes the objectives and characteristics of the questionnaire, and provides descriptive definitions of variables.
7. Manual del Encuestador	<i>Interviewer Manual</i> -- Provides guidelines for interviewers; describes fieldwork procedures, responsibilities of survey team members, time frame for assigned tasks, interviewing techniques, basic concepts and definitions, and objectives of each section; also, discusses how to identify the appropriate respondent, obtain the required information, and fill out the questionnaire.
8. Manual del Jefe de Brigada	<i>Team Leader Manual</i> -- Describes the responsibilities of the team leader: locating households, training and supervising interviewers, and checking questionnaire consistency.
9. Diccionario de Variables	<i>Dictionary of Variables</i> -- Lists all variable names by type (numerical, character, etc.), position, length and possible values. Variable names match those on the questionnaire, and are listed by section.
10. Manual del Codificación y Conversión	<i>Coding and Conversion Manual</i> -- Describes composition of key variables such as the household identifier (which can be broken down into segment, dwelling, and household number). Also lists codes for responses that are too numerous to list directly on questionnaire: geographic location, occupation, industrial activity, etc.
11. Data Dictionary - Community Questionnaire Data	<i>Community Questionnaire</i> -- dictionary of variables.
12. Means Tables	Provides means, standard deviations, and maximum and minimum values for all variables in household data files.

Note: These documents (with the exception of item 12 - "Means Tables") are available in Spanish only. To request copies, see Appendix A.

Questionnaires are pre-coded (coded responses are listed on the questionnaire, and are circled or marked by the interviewer), with the exception of a few questions for which possible responses are too numerous (name of village or town, industry of occupation, etc.). Instructions or notes to the interviewer appear on the questionnaire in capital letters. Questions read by the interviewer to the respondent appear in lower-case print. Capital letters enclosed in brackets within questions indicate where the interviewer should fill in the blank to complete the question (“How old is ...[NAME]...”). In each response field, the questionnaire includes a data identifier indicating what the response variable will be called in the data file.

For a detailed description of coding procedures and lists of codes, see the *Manual de Conversión y Codificación* (Coding and Conversion Manual).

The questionnaire is designed to allow the interviewer to skip easily over questions or sections that do not apply under given circumstances. An arrow followed by a number indicates the question number to which the interviewer should skip. If the interviewer should proceed to the following question, no instruction appears. Blank responses (questions that were omitted due to skip pattern) appear as blank fields in the data files.

The information from the household and community surveys can be analyzed jointly. For example, data on school attendance among individual children from the household questionnaire can be linked with data on the proximity and quality of schools as catalogued in the community questionnaire, in order to examine the relationship between attendance rates and access to schools.

2.1 The Household Questionnaire

The 1994 LSMS household questionnaire collects individual- as well as household-level information. The household questionnaire contains 14 sections covering different household characteristics and activities (see Table 2). The 1994 questionnaire contains several new sections or sub-sections vis-à-vis the 1991 questionnaire:

- **Section 4** (Health) contains a new sub-section on the health characteristics of women of child-bearing age;
- **Section 6** (Migration) contains a new sub-section on international migration;
- **Section 14** (Anthropometrics and Infant Health) is new.

Respondents vary by section; some sections require responses from each household member in order to gather data on individual characteristics (schooling, health problems, employment, etc.), while others collect information that pertains to the entire household (location and characteristics of the dwelling, durable goods owned, etc.) from the most informed household member with regard to that activity.

For the purposes of the survey, a household is defined as a person or collection of persons, whether related or not, that habitually live in the same dwelling, occupying it in part or in whole, and that care for life needs together. Members of the household are defined as those individuals that have habitually eaten and slept in the household for at least 3 of the 12 months prior to the interview date, with the following exceptions: (1) the household head (identified as such by the other household members) and infants under 3 months old are always considered members; (2) boarders and household employees are never considered household members.

The *Manual del Encuestador* (Interviewer's Manual) provides detailed definitions for concepts such as household, household member, etc.

Table 2: Sections of the 1994 Household Questionnaire

Section	Respondent
0. Household Identification	None
1. Household Demographics	Household head
2. Housing	Household head
3. Education	Household head
4. Health	All household members (parents may respond for children under 15)
5. Economic Activity	All members over age 5 (parents may respond for children under 15)
6. Migration	Only members 15 or older
7. Independent Employment	Most informed member regarding household-run business(es)
8. Non-Food Consumption Expenditures and Inventory of Durable Goods	Most informed member regarding non-food consumption
9. Food Consumption	Most informed member regarding food consumption
10. Other Income	Most informed member regarding income sources
11. Savings and Credit	Same as section 10
12. Location of Food Consumption	Same as section 9
13. Farm and Ranch Production	Most informed member regarding household activities in agriculture and animal husbandry
14. Anthropometrics	Children under 6 (mother also responds)

2.2 The Community Questionnaire

The community questionnaire was administered in rural communities by regional team leaders. Up to five community leaders were selected as respondents, with at least one female respondent per community. Common titles for respondents include: mayor, school principal, teacher, doctor, Mother's Club president, coordinator of the irrigation system, etc. The questionnaire covers demographics, economic activity (including a separate section for agriculture), access to education and health services, and local prices (see Table 3).

Table 3: Sections of the Community Questionnaire

Section	Description
0. Community Identification	Geographic location, name and title of respondents, date of interview.
1. Demography	Number of households, growth since 1991, gender mix of population.
2. Economy and Infrastructure	Primary economic activity of inhabitants, transportation and communication, migration.
3. Education	Primary and secondary schools accessible to community, year built, distance from community to school, time required to reach school, attendance of school-age children, reasons for non-attendance, literacy programs.
4. Health	Health services and practitioners accessible to community, hours of operation and distance from community, main health problems among population, main deficiencies of existing services.
5. Agriculture	Principal crops, technical assistance, cooperatives, daily wage for farm labor.
6. Prices	Market prices for 49 basic commodities.

3. Sample

3.1 The Household Survey Sample

The 1994 survey sample is designed to be representative of the population of Peru, based on census data from 1993. The sample frame consists of all private dwellings in Peru and their inhabitants.¹ The sample is stratified into seven geographic regions: Urban Coast, Rural Coast, Urban Sierra, Rural Sierra, Urban Selva, Rural Selva and Metropolitan Lima; results obtained can be broken down into these regions, and can also be grouped according to rural and urban areas.

¹ The sample frame excludes population groups that live in communal dwellings, such as army barracks, hotels, hospitals, asylums, monasteries, prisons, etc.

Urban areas are defined as those cities and towns with 2,000 or more inhabitants in the 1993 census.

Sampling was probabilistic and conducted in three stages.² The entire country was divided first into primary sampling units, then secondary sampling units (segments contain an average of 100 households), and finally into tertiary sampling units (individual dwellings).³ In the final stage of sampling, in all segments except for those in metropolitan Lima, 12 dwellings from each segment were selected to be interviewed for the survey. In Lima, 6 dwellings from each segment were selected. The entire sample consists of 3,544 dwellings, of which 820 were in metropolitan Lima, 1,380 in other urban areas, and 1,344 in rural areas.

Table 4: 1994 Sample -- Number of Segments and Dwellings, by Region

Region	Segments	Dwellings
<i>Metropolitan Lima</i>	<i>137</i>	<i>820</i>
<i>Other Urban Areas</i>	<i>115</i>	<i>1,380</i>
Coast	42	504
Sierra	41	492
Selva	32	384
<i>Rural Areas</i>	<i>112</i>	<i>1,344</i>
Coast	22	264
Sierra	60	720
Selva	30	360
<i>Total</i>	<i>364</i>	<i>3,544</i>

From the original sample of 3,544 dwellings, fieldwork resulted in a data set with observations for 3,623 households (recall that some dwellings contain more than one household). In some cases, it was not possible to complete interviews in the selected dwellings. For this reason, 422 dwellings originally selected for interview were replaced by other dwellings from the same segment. The known causes of replacement were refusal to participate (28.7 percent), unoccupied or abandoned dwellings (18.5 percent) absence of residents (16.1 percent), and failure to locate the dwelling (14.5 percent). Another 22.2 percent were replaced for unknown reasons.

For more information on the replacement of selected dwellings in the sample with alternates see *Información Básica* (Basic Information) which provides replacement statistics by region, and *Manual del Jefe de Brigada* (Team Leader Manual) on the procedures for replacing dwellings.

² In rural areas where primary sampling units had less than 500 inhabitants the sample was drawn in two stages.

3.2 The Community Survey Sample

The community survey was conducted in all rural areas covered by the household survey. The sample for the community survey covers all population centers within the geographic boundaries of the 112 rural segments included in the household survey sample. In many cases there was more than one population center within each household survey segment, such that a total of 204 population centers were interviewed.

For a list of communities (“centros poblados”) included in the community survey, see *Información Básica*, Annex 2. For coding of “centros poblados” see *Manual de Codificación y Conversión*, section 5.2.

3.3 Expansion Factors

Regional expansion factors must be applied to the data before the results can be aggregated and analyzed at the national level (within regions the sample is self-weighted). Each region is assigned a different weight to correct for regional variations in the number of households sampled relative to the population. Expansion factors also differ within region by rural and urban area (see Table 5).

Table 5: 1994 LSMS - Expansion Factors, by Region and Sample Segment

Region	Expansion Factor	Segments
Sierra (rural)	1,195.492	001-018, 031-051, 061-081
Sierra (urban)	1,441.866	100-102, 131-150, 161-178
Coast (rural)	740.532	424-425, 427-438, 472-479
Coast (urban)	1,641.710	103-129, 180-183, 461-471
Selva (rural)	685.636	605-611, 616-620, 625-627, 648-662
Selva (urban)	664.110	601-604, 612-615, 621-624, 628-647
Metropolitan Lima	1,503.401	235-362, 501-539

The expansion factors in Table 5 differ from those provided in the original survey documentation (see *Diseño Muestral*, p. 5). The new factors (in Table 5) are calculated based on the number of households that were actually interviewed in each region. The original expansion factors were based on the total number of households in the sample frame for each region.

³ It is important in Peru to distinguish between “dwelling” and “household.” A dwelling is a house, apartment or independent living space in which one or more households may live.

4. Survey Management and Fieldwork

The 1994 survey management team was headed by the Project Director (*Jefe de Proyecto*). The Director supervised all phases of survey work, both administrative and technical, and oversaw a team consisting of: (1) the Fieldwork Director (*Jefe de Operación de Campo*), responsible for training personnel, directing team of regional supervisors, and fieldwork execution; (2) the Administrative Director (*Jefe de Logística*), responsible for preparation and delivery of materials (questionnaires, maps), payroll, etc.; (3) the Systems Director (*Jefe de Cómputo*), responsible for data entry and consistency check.

Fieldwork took place between June and August of 1994. A brief (two-day) pilot test of the survey was carried out first; the test was short because the 1994 questionnaire contained only minor variations with respect to the 1991 questionnaire, which had functioned well. The pilot test focused on evaluating new sections of the questionnaire (which are listed in section 2.1 above).

Fieldwork was conducted by 15 teams, each consisting of 3 interviewers (*encuestadores*) and one team leader (*jefe de brigada*), and was directed by 5 regional supervisors (*supervisores regionales*). The team leader assigned each interviewer's workload; after each interview was complete, the team leader checked the questionnaire and if necessary returned it to the interviewer; the interviewer in turn returned to the household to verify or correct the data.

The household questionnaire required an average of two hours to complete. In most cases the interview was conducted in Spanish, but in the South and Central Sierra some interviews were conducted in Quechua or Aymara.

5. Deflation of Monetary Values

Values for aggregate household expenditure and income have been deflated to correct for inflation during the survey period.⁴

Monthly consumer price data provided by the Peruvian National Statistical Institute (*Instituto Nacional de Estadística e Información -- INEI*) for 16 major cities were used to deflate the survey data (see Appendix E). The data were deflated by assigning households (by segment) to the city that was closest in terms of geographic proximity or economic influence.

A second series of deflated monetary values is available for those variables used to calculate aggregate household expenditures, where regional differences in the cost-of-living are controlled for. For these alternative calculations, the price data collected in the community survey was used to estimate the average cost of a basic basket of consumption goods in each region. A cost-of-living index was then calculated as the ratio of the cost of the basket in each region

⁴ The national price level rose by approximately 3 percent during the survey period.

relative to its cost in Metropolitan Lima. Deflated monetary values were adjusted by this index, expressing each household's consumption expenditure in terms of the price level in Metropolitan Lima.

For more information on the cost-of-living index, composition of the basic consumption basket, and deflation procedures, see *Información Básica*.

6. Data Files

The data for each section of the questionnaire are stored in an individual file, or group of files. In addition to the original data files, files containing values for aggregate expenditures and income have been created and are available to users. Table 6 lists individual file names and describes their content.

Files containing expenditure (*gashog94*) and income (*inghog94*) aggregates are made available as part of the 1994 LSMS data for Peru. Basic information on the composition of these aggregates is available in Appendix D. For further details see *Especificaciones para la Composición del Ingreso y del Gasto* (Composition of Income and Expenditure Aggregates).

In each file, variable names begin with a different letter of the alphabet. This letter is used as a 'flag' to indicate the section (and page number, if sections contain more than one page) of the questionnaire to which the data file corresponds. For example, variables for section 0 (contained in file 'reg1') all begin with the letter 'A'; variables from section 1 begin with the letter 'B'; variables from section 2, page 1 begin with the letter 'C', and from section 2, page 2 with the letter 'D'; and so on.

Variable names combine this flag with the question number to which the variable refers. So for example in section 2, question 1 "What type of dwelling does the household live in?" is named "CO1"; question 2 "What is the primary material with which the dwelling is constructed?" is named "C02", etc. Variable names are printed directly on the questionnaire to facilitate their identification in the data files.

Each household has a unique identification number that can be created from the information contained in three variables: (1) segment (*segmento*) -- a three-digit number; (2) dwelling (*vivienda*) -- a two digit number ranging from 01-12 for originally selected dwellings and from 13-20 for replacement dwellings; and (3) household (*hogar*) -- a number ranging from

Table 6: Peru 1994 LSMS - Household Questionnaire Data Files

File Name	Questionnaire Section	Description
Reg01	Section 0	Household Identification
Reg02	Section 1	Characteristics of household members: age, gender, etc.
Reg03	Section 2A	Household Characteristics
Reg04	Section 2B	Household Expenditures
Reg05	Section 3-1	Education
Reg06	Section 3-2	“
Reg07	Section 3-3	“
Reg08	Section 4A-1,4A-2	Health
Reg09	Section 4B	Health of Women age 15 to 49
Reg10	Section 14-1, 14-2, 14-3	Anthropometrics: Infant Measurements and Basic Health
Reg11	Section 5A	Economic Activity of household members
Reg12	Section 5B-1	Primary Employment, past 2 weeks
Reg13	Section 5B-2	“ “
Reg14	Section 5B-3	“ “
Reg15	Section 5C-1	Secondary Employment, past 2 weeks
Reg16	Section 5C-2	“ “
Reg17	Section 5C-3	“ “
Reg18	Section 5D	Supplementary Job Search
Reg19	Section 5E-1	Primary Employment, past 12 months
Reg20	Section 5E-2	“ “
Reg21	Section 5E-3	“ “
Reg22	Section 5F-1	Secondary Employment, past 12 months
Reg23	Section 5F-2	“ “
Reg24	Section 5F-3	“ “
Reg25	Section 5G-1, 5G-2	Economic Activity, Cont'd
Reg26	Section 6A	Migration
Reg27	Section 6B	“
Reg28	Section 7A-1, 7A-2	Self Employment Activities (non-agricultural)
Reg29	Section 7BC	“ “ , Capital, Inventory, Expenditures
Reg30	Section 8A	Consumption: Daily Expenditures
Reg31	Section 8B	Consumption: Expenditures in Semi-durables and Services
Reg32	Section 8C	Consumption: Inventory of Durable Goods
Reg33	Section 8D	Expenditures for Transfers
Reg34	Section 12	Location of Food Consumption:
Reg35	Section 9A	Food Expenditures - past 15 days
Reg36	Section 9B	Consumption: value of home produced foods - past 15 days
Reg37	Section 10	Non-labor Income
Reg38	Section 11	Savings and Credit
Reg39	Section 11 cont'd.	“
Reg40	Section 13AB	Agriculture: land holdings and productions
Reg41	Section 13AB cont'd.	“
Reg42	Section 13CDE	Agriculture: by-products, forestry, equipment inventory
Reg43	Section 13CDE cont'd.	“
Reg44	Section 13CDE cont'd.	“
Reg45	Section 13FG	Agricultural inputs
Reg46	Section 13FG cont'd.	“
Reg47	Section 13HIJ	Livestock holding, production, by-products and expenditures
Reg48	Section 13HIJ cont'd.	“
Reg49	Section 13HIJ cont'd.	“
gashog94	various	Aggregate Household Expenditures
inghog94	“	Aggregate Household Income

Note: Monetary values in files 'gashog94' and 'inghog94' only have been deflated to regional prices of June 1994.

11-55 depending on the number of households in each dwelling.⁵ The household ID number can be created as follows:

$$\text{Household ID} = \text{segmento}(1000) + \text{vivienda}(10) + \text{int}(\text{hogar}/10)$$

Similarly, a unique identification number for each household member can be created by adding the variable “member number” to the household ID number, as follows:

$$\text{Personal Identification} = \text{segmento}(1000) + \text{vivienda}(10) + \text{int}(\text{hogar}/10) + \text{_00}$$

where “_00” is the variable “member number”. The generic name for this variable is “00”, however, in each data file it is preceded by the flag (see above) indicative of section and page number to which the information contained in that file corresponds. So for example, the flag for section 3, page 1 of the questionnaire is “e”. In this data file the variable ‘member number’ is named “e00”.

In addition to the files listed in Table 6, there are six additional files that users may find helpful (see Table 7).

Table 7: Peru 1994 LSMS -- Additional Miscellaneous Files

Filename	Questionnaire Section	Description
SYS8C	Section 8C	Corrections to selected variables and households
SYS13A	Section 13AB	Corrections to selected variables and households
SYS13H	Section 13HIJ	Corrections to selected variables and households
SYS13I	Section 13HIJ	Corrections to selected variables and households
INDICE1		Price indices for the 16 major cities (see also Appendix E)
UNIDSUPE	Section 13	Codes for surface units (unidades de superficie)

Files containing data from the community questionnaire are listed in Table 8.

Table 8: Peru 1994 LSMS -- Community Questionnaire Files

Filename	Questionnaire Section	Description
ccpp0	Section 0	Community Identification, Geographic Location
ccpp1	Section 1	Demographic Characteristics
ccpp2	Section 2: Questions 1 - 14	Economy and Infrastructure
ccpp3	Section 2: Questions 15 - 31	“
ccpp4	Section 3	Education
ccpp5	Section 4: Questions 1 - 3	Health
ccpp6	Section 4: Questions 4 - 9	“
ccpp7	Section 5: Questions 1 - 2	Agriculture
ccpp8	Section 5: Questions 3 - 13	“
ccpp9	Prices	Market Prices for Basic Goods

⁵ The variable *hogar* is a two digit number in which the first digit indicates the ordinal number corresponding to the household in the dwelling and the second digit indicates the total number of households in the dwelling.

7. Subset of Panel Data -- 1991 and 1994 Data Sets

The LSMS surveys for Peru all contain a subset of panel observations -- households that have been interviewed in more than one survey year. This section contains the available information on identifying panel observations in the 1991 and 1994 survey samples. A file containing this subset of observations is not currently available from the World Bank.

The 1994 LSMS sample includes all 2,522 dwellings where interviews were conducted in 1991, representing 63.5 percent of the total sample for 1994. The number of actual panel observations will be substantially less as households will have moved, significantly changed in composition, denied the interview, etc.

Because the 1991 sample was not representative of all regions of the country, the subset of panel observations for 1991-1994 also do not represent all regions. Analysts working with the panel may wish to consult *Basic Information: Peru -- Living Standards Measurement Survey (PLSS) 1991* for a more detailed description of the 1991 sample and its limitations. Expansion factors to correct the 1991 data for the under-representation of certain regions have also been developed (see Appendix B).

In the 1994 data set, variable A16, section 0, identifies 1,448 dwellings where interviews were also conducted in 1991. The data for these panel households in the 1991 data set can be identified initially according to the household ID number, which under normal circumstances will not have changed between the two survey years. Only in cases where there was more than one household within a single dwelling is it possible for the household ID to differ by survey year (difference will occur in the last digit only).⁶ Initial work with the panel indicates that of the 1,448 households identified as panel observations according to variable A16, 1,222 have the same household ID number and name of household head.

Household size and composition will have changed among panel households. In the 1994 survey, variable B11, section 1 identifies 5,833 individuals that were reportedly interviewed in 1991 as well. Matches may be confirmed using other information collected in section 1 -- age, sex, and date of birth. Initial work with the panel was able to match only 3,170 of these individuals with observations in the 1991 data.

Work with the above panel data will require expansion factors for 1991, and price indices allowing monetary values across surveys to be compared in real terms. Appendix B provides revised expansion factors for the 1991 data, and Appendix C contains information for comparing monetary values across the 1991 and 1994 surveys.

⁶ The ordering of households by number within a dwelling may have varied, or new households may have formed within a dwelling between the survey years, changing the numbering pattern.

8. Data Quality

As a whole the 1994 LSMS provides high quality data, given the close supervision of fieldwork, and both computerized and manual consistency checks of completed questionnaires. This section highlights those errors and particularities that have been brought to light in the course of the World Bank's work with the data to date.

1) Data Entry Errors

- a. In segmento 78, vivienda 8 the dwelling identified as hogar 14 is actually hogar 11.
- b. In segmento 653, vivienda 7, hogar 11, the value for variable L02 should be 113, not 13 as currently recorded.

2) Duplicates

Some of the data files appear to contain duplicate records, where information is repeated at the individual or household level. Users should be careful to check for the possibility of duplicated information prior to running analyses.

3) Identifying Intra-family Relationships

All of the Peru LSMS surveys contain a weakness in questionnaire design that complicates the identification of relationships between household members. The household roster (section 1) positively indicates each member's relationship to the head, but not to other household members. For example, while the spouse of the head of household is positively identified as such, spouses of other members living in the household are not clearly identified. However, entries on the roster follow a strict order indicating their relationship to each other. For example, after the head of household and partner, their single children are listed first in order of age, then each married child is followed by his or her partner and their children, etc. As a result, some programming will be required in order to link married partners that are not household heads, brother-in-laws with sisters, grandmothers with grandfathers, etc., and in some cases relationships may not be clearly discernible.

The *Manual del Encuestador* (Interviewer Manual) provides complete specifications on the order in which household members appear on the roster.

4) Imputed Rent

Analysis of the 1994 household expenditure data uncovered a disproportionately high share of total expenditures attributed to household rent. Whereas in 1991 rent (both imputed and real) represented 13 percent of total household expenditures, in 1994 this share had risen to 24 percent. The source of this discrepancy was identified as imputed rent - the estimated rental value of housing currently owned and lived in. As a result, in the existing aggregate expenditure files

the value of imputed rent for any given household is capped at 30 percent of its total expenditures. Imposing this restriction reduces the average share of rent in total expenditures to 16 percent. The calculations of imputed rent, and the relationship between the current capped rents and household expenditures may warrant further study. The relevant variables (in data file *gashogli*) are:

M34AJ = imputed rent capped at 30 percent of household expenditures.

M34 = imputed rent.

5) Codes for Industrial Activity

The codes used to classify industry activity have changed between 1991 and 1994. The new codes for industrial activity are listed in the *Manual del Codificación y Conversión*; for codes used in 1991 and prior years, see *Censos Nacionales: VIII de Población y III de Vivienda, Anexo 04: Código de Ocupaciones* (to request a copy, see Appendix A).

For information on the calculation of imputed rent, see *Especificaciones Para la Composición del Ingreso y del Gasto* (Composition of Income and Expenditure Aggregates), section 1.6.

Appendix A: How to Obtain LSMS Survey Data and Documentation

The 1994 Peru Living Standards Survey Data are the property of the World Bank and Cuánto, S.A. Data and supporting documentation are available free of charge on the World Bank's LSMS world wide web site. The address for the home page is:

<http://www.worldbank.org/lms/lms/home.html>

Users who do not have access to the world wide web may send requests for data and documentation to the following address, along with a brief description of their intended research:

Living Standards Measurement Study (LSMS)
Development Research Group
Poverty and Human Resources
The World Bank
1818 H Street North West
Washington, D.C. 20433
U.S.A.

email: LSMS@worldbank.org

A nominal fee is charged for copying and mailing the data files, which the World Bank will provide on 3 1/2" diskettes in SAS portable (XPORT engine), Stata (version 2.1) or ASCII formats. Users should indicate their preferred format. The Poverty and Human Resources Division of the World Bank requests copies of all reports and documents resulting from research with the data.

Once received, the data cannot be passed on to a third party for any reason. Other researchers must contact 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 B: Revised Expansion Factors for the 1991 Data

The 1991 survey sample was designed based on population estimates for that year calculated from the most recent census data available at the time - 1981. Initial expansion factors (column 1, Table B.1) expand the survey sample population to represent the distribution of the 1991 population as projected from the 1981 census. The 1993 census data allow these projections to be refined and the 1991 expansion factors to be revised accordingly (column 2, Table B.1).

These revised expansion factors may be used for 'stand alone' work with the 1991 survey. A second set of revised figures have been developed for use in comparing the 1991 and 1994 survey results (column 3, Table B.1). The 1991 LSMS was conducted in a time of terrorism in Peru, omitting unstable areas within some regions (Ayacucho and Huancavelica from the Central Sierra; Apurimac from the South Sierra), and excluding other regions in full (the Central Coast and all of the Selva). The second set of expansion factors can be used to expand the areas surveyed in 1991 to represent the entire regional population in that year. Use of these factors requires the assumption that the characteristics of the surveyed areas are representative of those excluded; on the other hand, using the first set of revised factors (that expand only to the population of the surveyed areas) will under-estimate regional populations in 1991 relative to 1994.

Table B.1: 1991 Expansion Factors by Region, Initial and Revised

Table B.1: 1991 Expansion Factors by Region, Initial and Revised			
Region	Initial	Revised	
		<u>Surveyed Areas^a</u>	<u>Entire Region^b</u>
<u>Rural</u>			
North Sierra	1497.57	1635.4	1635.4
Central Sierra	854.75	879.7	1310.5
South Sierra	1265.74	1400.1	1679.9
<u>Urban</u>			
North Sierra	1002.93	1263.7	1263.7
Central Sierra	1175.05	1036	1318.7
South Sierra	1376.2	1273.2	1380.4
North Coast	1627.93	1550.5	2079.2
South Coast	1103.69	1071.4	2079.2
Metropolitan Lima	1521.55	1425.4	1425.4

^a Refers only to those areas surveyed. This excludes Ayacucho, Huancavelica from the Central Sierra; Apurimac from the South Sierra; and the entire Selva and Central Coast.

^b Refers to all areas in each region.

Appendix C: Price Indices to Compare Monetary Values in 1991 and 1994

There are three ways to bring the 1991 data files, which are expressed in prices of Metropolitan Lima, to values comparable with similar files for 1994 (Nuevo Soles of Lima of June 1994):

- 1) change in the price of the Lima food basket (as calculated by Cuánto) = 187 percent
- 2) change in the price of the Lima food + non-food basket (as calculated by Cuánto) = 190 percent
- 3) change in the Lima price index (as calculated by INEI) = 163.77 percent. Here the prices indices are monthly averages and so an average of September and October is taken to get a value representing the first week of October (October 1 = 616.1665 and Mid-June = 1625.251).

Appendix D: Composition of Household Income and Expenditure Aggregates

Aggregate values for household income and expenditures have been calculated by Cuánto, S.A. and are available as part of the 1994 Peru LSMS data files. This appendix describes the composition of these aggregates. Further information, including exact calculation methods, may be obtained from *Especificaciones para la Composición del Ingreso y del Gasto*.

Household Income

Total household income is calculated as the sum of:

- 1) Each member's earnings from wage employment and self-employment, including:
 - a) base wages;
 - b) additions to base wages (given on a weekly or longer interval);
 - c) income in-kind for employer-provided goods and services such as meals, transportation, etc;
 - d) self-employment income;
 - e) farm labor.

Note: There is no explicit treatment of income from farming as a second occupation in the 1991 survey; comparisons of employment income between the 1991 and 1994 surveys will need take account of the addition of this income component in 1994.

- 2) Estimated value of goods produced and consumed by the household.
- 3) Incomes from property rent.
- 4) Regular payments received by the household (pensions, dividends, transfers from family members, etc.).
- 5) Irregular payments received (lottery winnings, inheritances, etc.).

Household Expenditures (Consumption)

Total household expenditures on consumption goods is calculated as the sum of:

- 1) Foods consumed, including:
 - a) food, drink, and tobacco purchases;
 - b) estimated value of self-produced food, drink and tobacco;
 - c) estimated value of food received as in-kind payments from employer (category actually includes transportation services received as well as food items).

- 2) Clothing purchased and received in-kind from employer.
- 3) Housing expenses, including:
 - a) rent;
 - b) utilities and local property taxes;
 - c) other regular expenses.
- 4) Household maintenance expenses, including:
 - a) purchases of furniture and appliances;
 - b) purchases of products for household cleaning and repairs;
 - c) maid services;
 - d) estimated value of goods produced by own business and used for household consumption.
- 5) Expenses for health care and medicine.
- 6) Expenses for transportation and communications, including:
 - a) public transportation, gasoline, and mail;
 - b) maintenance and repair of vehicles;
 - c) long-distance travel;
 - d) purchase of vehicle for personal use (automobile, motorcycle, bicycle);
 - e) telephone service.
- 7) Education and entertainment expenses, including:
 - a) magazines and newspapers;
 - b) books, games, other entertainment items;
 - c) purchase of radio, television, camera, etc.
 - d) schooling: tuition, transport to school, meals, school supplies;
 - e) pre-school expenses.
- 8) Other goods and services, including:
 - a) meals purchased at restaurants;
 - b) goods for personal hygiene;
 - c) other products produced by home business and consumed by household.
- 9) Value of transfers (monetary or in-kind) given to non-household members.

Appendix E: Deflators used for 1994 ENNIV

Table E.1 Deflators Used in ENNIV 1994 to Give Monetary Values to Prices in Each Domino and to June 1994

City	1993				1994							
	Sept	Oct	Nov	Dec	Jan	Feb	Mar	April	May	June	July	Aug
Abancay	1.19243	1.16609	1.14521	1.11563	1.09240	1.07232	1.04470	1.02307	1.01409	1.00000	0.98973	0.97214
Arequipa	1.16581	1.14740	1.12762	1.09845	1.07714	1.05890	1.04014	1.02769	1.01336	1.00000	0.98902	0.97528
Cajamarca	1.15960	1.14355	1.12756	1.10242	1.08269	1.06328	1.03768	1.02052	1.01280	1.00000	0.98844	0.97518
Chiclayo	1.15450	1.13566	1.12000	1.09434	1.07397	1.05468	1.02924	1.01292	1.00703	1.00000	0.98867	0.97248
Chimbote	1.18271	1.09985	1.13840	1.11624	1.09640	1.07503	1.04588	1.02316	1.01209	1.00000	0.98641	0.96749
Cusco	1.18433	1.16306	1.13837	1.10438	1.08135	1.05962	1.03330	1.01870	1.01206	1.00000	0.99068	0.99068
Huancayo	1.16595	1.14927	1.13462	1.10556	1.08484	1.06468	1.04185	1.02566	1.01310	1.00000	0.98484	0.97352
Huánuco	1.17522	1.15655	1.12981	1.10475	1.08463	1.06386	1.03931	1.01798	1.01163	1.00000	0.98694	0.96905
Ica	1.16113	1.14207	1.12327	1.09645	1.07790	1.05949	1.03782	1.02035	1.01399	1.00000	0.99193	0.97318
Piura	1.16823	1.14882	1.13237	1.10266	1.08224	1.06393	1.04173	1.02336	1.01380	1.00000	0.98704	0.97093
Puerto	1.18141	1.15319	1.13414	1.10612	1.08449	1.06560	1.04252	1.02427	1.01365	1.00000	0.98483	0.96598
Puno	1.17226	1.15142	1.13023	1.09981	1.07775	1.05612	1.03216	1.01966	1.01038	1.00000	0.98799	0.97440
Tacna	1.16746	1.14903	1.12855	1.09913	1.08243	1.06377	1.03653	1.02014	1.01184	1.00000	0.98576	0.97608
Trujillo	1.17425	1.15310	1.13449	1.10792	1.08719	1.07072	1.04336	1.02555	1.01489	1.00000	0.98508	0.96714
Lima	1.16035	1.14311	1.12507	1.09750	1.07769	1.05842	1.03438	1.01865	1.01141	1.00000	0.99118	0.97621
Iquitos	1.17051	1.14675	1.12127	1.08431	1.07063	1.05350	1.03105	1.01494	1.00885	1.00000	0.98877	0.97354

This table is Table 14 from the Basic Information Document.

These values are also found in file INDICE1.