

Vietnam Living Standards Survey (VLSS), 1997-98

Basic Information

Poverty and Human Resources Division

The World Bank

April 2001

PRINCIPAL ABBREVIATIONS AND ACRONYMS USED

GSO	General Statistical Office
LSMS	Living Standards Measurement Study
MPHS	Multi-Purpose Household Survey
MPI	Ministry of Planning and Investment
PSU	Primary Sampling Unit
SIDA	Swedish International Development Authority
SPC	State Planning Committee
UNDP	United Nations Development Program
VLSS	Vietnam Living Standards Survey

This document was prepared as part of an expanded program of documentation and further development of the Living Standards Measurement Study (LSMS), managed by Kinnon Scott, in the Poverty Division of the Development Research Group (DECRG). It was prepared by Sarah Bales (consultant) with the assistance of the Vietnamese General Statistical Office.

CONTENTS

1	Overview.....	1
2	Survey Questionnaires	1
2.1	Household Questionnaire.....	1
	First Interview.....	3
	Second Interview	12
2.2	Commune/Ward Questionnaire	18
2.3	School Questionnaire.....	20
2.4	Commune Health Station Questionnaire	21
2.5	Price Questionnaire.....	21
3	Sample	22
3.1	Sample Design	22
3.2	Implementation.....	23
3.3	Weights or Expansion Factors	24
3.4	Complications in calculating weights (expansion factors)	26
	Missing Data problem.....	26
	Incorrect formula problem.....	27
	Evaluation of Weights (Expansion factors)	28
	Sources of information for calculating weights	29
3.5	Selection of respondent for the Section on Fertility	30
4	Organization of the Survey	30
4.1	Survey Management	30
4.2	Training and Field Testing.....	31
4.3	Organization of the Field Work	32
	Field work.....	32
	Timing.....	33
5	Using the Data	33
5.1	Data Processing	34
5.2	Linking Parts of the Household Questionnaire.....	35
5.3	Linking the Household to the Price and Community Data	36
6	Constructed Data Sets.....	36
6.1	Aggregate Expenditure	37
6.2	GIS Information.....	37
	Appendix A. How to Obtain the VLSS 1997-98 Data set	38
	Appendix B. List of VLSS Data Sets Available	39
	Appendix C. List of Related Documents	42
	Appendix D. Price Indices	43
	Appendix E. Household Expenditure Calculations	47
	Appendix F. Region Codes	56
	Appendix G. List of Communes, Districts and Provinces Sampled	59
	Appendix H. Occupational and Industrial Classifications Used in the VLSS 97-98.....	64
	Appendix I. List of Reports/papers Using the VLSS 1997-98	65

Vietnam Living Standards Survey (VLSS), 1997-98

Basic Information

1 Overview

The first Vietnam Living Standards Survey (VLSS) was conducted in 1992-93 by the State Planning Committee (SPC) (now Ministry of Planning and Investment) along with the General Statistical Office (GSO). The second VLSS was conducted by the GSO in 1997-98. Both VLSS surveys were funded by UNDP and Swedish International Development Authority (SIDA). The survey was part of the Living Standards Measurement Study (LSMS) household surveys conducted in a number of developing countries with technical assistance from the World Bank.

The second VLSS was designed to provide an up-to-date source of data on households to be used in policy design, monitoring of living standards and evaluation of policies and programs. The timing of the second VLSS approximately five years after the first allows analysis of medium term trends in living standards as a large part of the questionnaire is the same in both surveys.

In addition to the purpose of obtaining a comprehensive and comparable data set to the 1992-93 VLSS for policy analysis, the survey also served as a medium for training and improving survey methods and analysis within the General Statistical Office of Vietnam (GSO), the agency in charge of designing and implementing the second round of the VLSS as well as other government agencies involved in social statistics.

This document provides background information on the second round of the Vietnam Living Standards Survey and the data-sets resulting from this nation-wide household survey conducted between December 1997 and December 1998. Information is provided on the survey design and implementation, the content of the different questionnaires, data processing activities, the different data-sets available and how to obtain permission to use the data.

2 Survey Questionnaires

The second round of the VLSS used 5 questionnaires: household, commune, price, school and clinic.

2.1 Household Questionnaire

The household questionnaire contains 15 sections each of which covered a separate aspect of household activity.

SURVEY INFORMATION

0. HOUSEHOLD ROSTER
1. EDUCATION
2. HEALTH
3. EMPLOYMENT
4. MIGRATION
5. HOUSING
6. RESPONDENTS FOR SECOND ROUND
7. FERTILITY
8. AGRICULTURAL, FORESTRY AND FISHERY ACTIVITIES
9. NON-FARM SELF-EMPLOYMENT
10. FOOD EXPENSES AND HOME PRODUCTION

11. NON-FOOD EXPENDITURES AND DURABLE GOODS

12. INCOME FROM REMITTANCES

13. BORROWING, LENDING AND SAVINGS

14. HEALTH CHECK (height, weight and arm circumference)

For some sections (survey information, housing and respondents for the second round), the individual designated by the household members as the household head provided responses. For some others (agro-pastoral activities, non-farm self-employment, food expenditures, non-food expenditures) a member identified as most knowledgeable provided responses. Identification codes for respondents of different sections indicate who provided the information. In sections where the information collected pertains to individuals (education, health, employment, migration and fertility) each member of the household was asked to respond for himself or herself. Table 1 indicates the share of respondents who responded for themselves for each section. Most of those who did not respond directly were children, but there are a few cases of people who were not available to be interviewed despite repeated visits by interviewers.

There were numerous changes in the questionnaire compared to the 1992-93 VLSS. In particular, in the first half, two pages are added to aid in linking data on the panel households, section 2 and 3 were revised substantially and agricultural labor hours were asked separately from other employment in a format that is more appropriate for the irregular work schedule of farming activities. In the second half, the section on agricultural land was thoroughly revised, and a section on household lending was added. Other minor changes compared to the 1992-93 VLSS will be noted in each section. Note that in 1997-98, all codes for responses are printed in the questionnaire itself.

Table 2.1: Respondent rates for Various sections of the VLSS 1997-98

Section	Respond for self	Relevant respondents
Education	70%	All household members
Health	70%	All household members
Employment	76%	All household members age 6 and above
Fertility	99%	A randomly selected ever-married woman aged 15-49 in the household
Farming	99%	The person who knows the most about agricultural activities of the household
Self-employed non-farm	98%	Person who knows the most about each self-employed non-farm activity
Food expenditures	99%	Person who knows the most about food expenditures of the household
Non-food expenditures	99%	Person who knows the most about non-food expenditures of the household

The household questionnaire was completed in two interviews two weeks apart: Sections 0-8 were conducted in the first interview, sections 9-14 were conducted in the second interview, and section 15 was administered during both rounds, whenever household members could be found at home.

The survey was designed so that more sensitive issues such as credit and savings were discussed near the end. The content of each module is briefly described below.

First Interview

Section 0 SURVEY INFORMATION

0A HOUSEHOLD INFORMATION AND DATES OF INTERVIEWS

0B STATUS OF 1992-93 DWELLING AND HOUSEHOLD

0C SUMMARY OF SURVEY RESULTS

0D OBSERVATIONS AND COMMENTS

The date of the interview, the religion, ethnic group of the household, language used by the respondent and other technical information related to the interview are noted. Section 0B provides information on the status of dwelling and household interviewed in 1992-93 along with the cluster and household identification codes to link this data with that from 1992-93. Section 0C summarizes the results of the survey visits, i.e. whether a section was completed on the first visit or the second visit. Section 0D provides in string form various comments on the reliability of household information, specifies information when the “other” code was used, and provides other information on unusual cases.

Differences with the 1992-93 questionnaire

Section 0A Codes now exist for all ethnic groups, In 1997-98 a more precise definition of religion was applied requiring that the household be practicing the religion, not just reporting that they identify themselves with a particular religion.

Section 0B is completely new to provide information on panel households

Section 0C is reformatted to collect information on each visit to the household rather than by section

Section 0D comments from interviewers are entered in the computer files in Vietnamese. These include specifications of “other” responses for interviewers who closely followed instructions.

For households not interviewed in 1997-98 who were in the 1992-93 VLSS, information is only collected on sections 0A, 0B and 1D. Information is available on 373 such households.

Section 1 HOUSEHOLD MEMBERSHIP

1A HOUSEHOLD ROSTER

1B INFORMATION ON PARENTS OF HOUSEHOLD MEMBERS

1C CHILDREN OF HOUSEHOLD MEMBERS WHO ARE LIVING ELSEWHERE
AND ARE NOT MEMBERS OF THE HOUSEHOLD

1D MEMBERS OF THE HOUSEHOLD IN VLSS 1992-93

The roster in section 1A lists the sex, relation to head of household, age, birth dates, marital status of all people who spent the previous night in that household and for household members who are temporarily away from home. The household head is listed first and receives the personal id code 1. Household members were defined generally to include “all people who normally live and eat their meals together in this house and have done so for 6 or more months out of the past year” which is the same as in 1992-93. However, specific cases to include as members or exclude as non-members differ slightly from 1992-93 and are listed in the questionnaire and in Table 2:

Table 2.2: Categories of household members and Non-members

Members	Non-Members
Household head	Individuals who died during the past 12 months
1997-98-Relatives of household head absent 6 months or less in the past 12 months who are not permanent residents of other households 1992-93 relatives of household head absent 6 months or less in the past 12 months	1992-93 People who have lived in the household more than 6 months, but left due to marriage, etc. 1997-98 People who have lived in the household more than 6 months, but left due to marriage, etc. and are no longer permanent household members
Infant less than 6 months old	1992-93 Hired workers, servants, or lodgers if they are members of households elsewhere 1997-98 Hired workers, servants, or lodgers
New permanent residents of the household less than 6 months because they were newly demobilized, married, or had a job transfer.	Guests and all other people not listed in the definition of household members
Students living outside the household but still supported by their family and not members of other households	1997-98 People who have joined the army but plan to return to this household (originally member in questionnaire, but in professional notice number 5 on 8 April, 1998, they were reclassified as non-members)
1997-98 Relatives of household head whose work requires them to be outside the household more than 6 months of the year, but who consider this household their permanent home and contribute to the household budget	
1992-93 Guests living with the household 6 or more months	

Note: If year is stated, the category only applies to that year.

A lunar calendar is provided in the questionnaire to help respondents recall the year and month they were born. For individuals who are married and whose spouse resides in the household, the personal id number of the spouse is noted. This way information on the spouse can be collected by appropriately merging information from the roster and other parts of the survey.

Section 1B collects information on the parents of all household members aged 70 or less. For individuals whose parents reside in the household, parents' personal identification numbers are noted, and information can be obtained by appropriately merging information from other parts of the survey. For individuals whose parents do not reside in the household, information is recorded on whether each parents is alive, as well as their schooling and occupation.

In Section 1C information is collected for all children of household members living elsewhere. No age restriction is made on which children to include. Children who have died are not

included. All living children are listed along with the personal id number of their father or mother (if parents reside in the household), age, educational attainment and current place of residence.

Section 1D is designed to provide information on the current whereabouts of household members who participated in the 1992-93 survey. Information on name, personal id code, sex, and age were copied from the household roster of the 1992-93 VLSS questionnaires. During the interview in 1997-98, information is collected on whether the individual is still a household member, their current ID, or death date, or current place of residence. For households not interviewed in 1997-98 who participated in 1992-93, attempts were made to collect information in 1D through neighbors, relatives or local authorities.

Differences with the 1992-93 Questionnaire

Section 1A In 1997-98, information is not asked about the nationality of the individual as it was in 1992-93. Adoptive children were separated from biological children in the relationship codes and placed with step-children. In-law categories were dropped. Servants and other workers were combined into the non-relative category.

Section 1B In 1992-93 this section included even those above 70. The codes for school attainment are more detailed in 1997-98 than in 1992-93. In 1997-98 the codes for major occupation of parents is listed in the questionnaire.

Section 1C In 1992-93 this section only included children under 30. In 1997-98 the question on whether the current place of residence is urban or rural was dropped. The 1997-98 codes on education attainment are more detailed than in 1992-93. The codes for provinces have changed as several provinces have split between the first and second implementation of the VLSS.

Section 1D is entirely new.

Section 2 SCHOOLING

2A INFORMATION ON PREVIOUS SCHOOLING

2B INFORMATION ON CURRENT SCHOOLING

2C REPEATING GRADES

2D EDUCATION EXPENDITURES

2E LITERACY AND VOCATIONAL/TECHNICAL TRAINING

Section 2 is completely revised compared to the 1992-93 survey so comparability is a problem, especially for literacy and for educational attainment variables.

As Vietnam has had many different educational systems over the past century, a table of education equivalents is presented in the 1997-98 questionnaire and the interviewers are asked to record current equivalent years of schooling rather than the reported years of schooling. For example, until recently there were only 10 grades in general education in the north. For people who attended that school system, instead of recording the highest number of years of general schooling as 10, the interviewers would record 12 which is the current highest grade of general education throughout the country.

Section 2A first sorts out whether an individual has ever been to school, and if yes, whether they are currently in school or have completed their schooling. Depending on the answer, the interview continues in section 2A, 2B or 2E.

Section 2A then goes on to ask about the last level of schooling attended, years completed, degrees obtained, university subject of study, and type of school and date when stopped going to school.

Section 2B asks about school attainment for those currently in school. This includes questions on whether living at home while in school, code of the school to link with school questionnaire, time in school in past week, current grade level, type of school (private/public), grades in past examinations and subject of university study.

Section 2C asks about grade attained prior to attending technical training, whether the individual repeated any years, and how many years repeated at each level of schooling. For those under 20 years of age, the year and age of entry into primary school is asked.

Section 2D asks details about school expenditures including tuition, private tutoring, parent's association, construction fund, exam papers, special events, uniforms, textbooks, school supplies, transportation and lunch money, other expenses and a total. Individuals were asked whether or not they were allowed exemptions or reductions in fees, and for what reason. Students were also asked about the value of scholarships and whether or not they had sufficient textbooks.

Section 2E asks about literacy and vocational training. The respondent was asked about apprenticeships or internships, then about vocational/technical training. They were asked for up to two training programs whether training was completed, what level of the training system, how long the training lasted, which field of training and whether received certificate. The final questions asked about study of foreign languages or computers and the cost of this training, or any other training obtained in the past 12 months.

Differences with the 1992-93 Questionnaire

In 1997-98 all household members were interviewed regarding schooling while the 1992-93 VLSS only asked those preschool age and higher.

In 1997-98 educational attainment is asked separately for those currently in school and those who have completed their schooling while in 1992-93, the same questions were asked of both cases. In 1997-98 more detailed information about technical training is asked.

Section 2A and 2B Questions on educational attainment in 1997-98 ask whether people have completed an educational level and received the diploma, completed and not obtained the diploma or not completed. In 1992-93 they were only asked the total number of years completed and highest degree obtained. 1997-98 Codes for educational attainment separate out junior college, 2-year college certificate and bachelor's degree which were all coded as college/university in 1992-93. Subject codes in 1997-98 are more detailed than in 1992-93. In 1997-98 more than one university degree could be reported. Codes on ownership type of school in 1997-98 separate private and community sponsored. In 1997-98 the name of the school was asked and can be linked to information in the school questionnaire.

Section 2C is completely new asking about repeating years and age and year at entry to school

Section 2D In 1992-93 students were not asked specifically about costs of private tutoring, special events, construction funds, exam fees, textbook rental or purchase as these were included in the "other" category. In 1997-98 students were not only asked about scholarships, but also about exemptions or reductions in school fees and for what reason. When asking the amount of the scholarship in 1997-98, subsidies are included while in 1992-93 only scholarships are included. In 1992-93 distance to school was asked, but in 1997-98 that was only asked at the village level in the commune questionnaire but can be linked to the individual through the school code in question 4 of Section 2B.

Differences with the 1992-93 Questionnaire

Section 2E Assessment of literacy in 1997-98 was made by asking respondents whether they could read a short passage printed in the questionnaire and the interviewer making an assessment of whether the passage was read easily, with difficulty or not at all. Numeracy was evaluated in a similar way by asking the respondent to do calculations printed in the questionnaire. In 1992-93 respondents were asked whether they could read and write, read only or neither. In 1992-93 assessment of numeracy was also based on self-reporting of individuals.

In 1992-93 respondents were only asked about apprenticeships, not internships.

In 1992-93 respondents were asked about whether they had a technical worker or technical high school degree while in 1997-98 they were asked about post-primary, post-lower secondary, professional high school, technical high school or other. It is difficult to make equivalent categories between the two surveys. The training subject codes are more detailed in 1997-98.

In 1992-93 respondents were only asked about foreign language training, while in 1997-98 they were also asked about computer training. In 1997-98 the language of training was not asked. In 1997-98 a question was asked about any further training expenditures beyond language, but in 1992-93 this was not asked.

Section 3 HEALTH

Section 3 was substantially revised and improved providing much more detail than in the 1992-93 VLSS.

The section begins by asking respondents whether or not they have had any of a list of 8 symptoms in the past 4 weeks, or whether they suffered from injury. If ill in the past 4 weeks, they are asked about number of days ill, and number of days unable to do normal activities because of illness.

A series of questions on health seeking and total costs, cost of medicine and cost of transport, room and board over a 4 week reference period is asked for various types of providers including Public hospital, Commune health station, Regional polyclinic, Other State health facility, Private western practitioner, and traditional practitioner. Questions are asked about house visits by practitioners and visits directly to pharmacists over the past 4 weeks including the amount spent on these uses of health care and transport to get there.

In addition, questions are asked about number of nights spent in a hospital over the past 12 months, the cost of those hospital visits, and an estimate of total amount spent on all health care excluding those hospital visits in the past 12 months. Individuals are asked whether they have health insurance and the cost of that health insurance.

A series of questions on use of tobacco products is asked of all individuals age 6 and over including whether they have ever smoked more than 6 months, whether they currently smoke, how many cigarettes, or how much other tobacco products they use and how much it costs them.

Differences with the 1992-93 Questionnaire

In 1992-93 one question was asked on whether the respondent was ill or suffered from injury in the past 4 weeks, then in the past 12 months. While examples of illness included cough, cold, diarrhea and accident were included in the question, the response was much lower than in the 1997-98 survey which asked specific questions about each symptom. In 1992-93 respondents were then asked to self report which illness they suffered from (only one). In 1992-93 but not 1997-98 they were asked about how long ago this illness or injury occurred.

In 1992-93 respondents were only allowed to respond about one caregiver, one health care facility, and whether it was private or public, and whether they used western or traditional medicine or a combination. Respondents were asked about the distance to the caregiver from home, while in 1997-98 this question was asked in the community questionnaire. In 1992-93 only one question was asked about cost of transport to health care provider listed and the cost of consultation. A question is asked about the cost of medicines used in the past 4 weeks in general, not specifying provider.

In 1992-93 a 4 week reference period is used for in-patient hospital visits, compared to 12 week reference period in 1997-98.

In 1997-98, the questions on tobacco use ask about cigarettes, pipe tobacco, chewing tobacco, and self-rolled cigarette tobacco (used in the south), while in 1992-93 they were asked about cigarettes, pipe and chewing tobacco but not the self-rolled cigarette tobacco. In 1997-98 each individual was asked about the amount spent on tobacco products, while in 1992-93 they were only asked about the amount of cigarettes and tobacco, while a household level question was asked about tobacco expenditures.

Section 4 EMPLOYMENT

4A TYPE OF WORK AND JOB SEARCH

4B MAIN JOB DURING PAST 7 DAYS

4C SECONDARY JOB DURING PAST 7 DAYS

4D MAIN JOB IN PAST 12 MONTHS

4E SECONDARY JOB IN PAST 12 MONTHS

4F AGRICULTURAL WORK IN PAST 12 MONTHS

4G EMPLOYMENT HISTORY

4H OTHER ACTIVITIES

All individuals age six and older were asked to respond to the economic activity questions in Section 4, beginning with questions on the nature of their work in the last seven days. For persons who did not work in the past seven days, data were collected on job search and reason for not seeking employment. For work in the last seven days, information was collected on hours, length of employment, type of employer, distance and travel time to place of work, money and in-kind compensation and benefits. Similar questions were asked on the secondary job in the last seven days. If main work in the past twelve months was different from the main or secondary job in the past seven days, the complete set of questions was answered for that work as well. Likewise for a secondary job in the past 12 months. For those in self-employed agricultural work, a different series of questions was asked on hours worked in peak and non-peak weeks in the past 12 months for six different agricultural-related work activities. Type of work and years of experience at any work prior to that of the main job in the last twelve months were collected. Days and hours spent doing household chores were collected for each household member age 6 and older. Occupation and industry of employment codes are printed directly in the household questionnaire.

Differences with the 1992-93 Questionnaire

Section 4A In 1992-93 a question is asked about whether the individual seeks wage or self-employed work, and the methods they are using to seek work but not in 1997-98. In 1997-98 an additional question is asked if not working and not seeking work to determine whether the respondent has a regular job and if so, why they are not working in the past 7 days. Respondent codes on reasons for not working differ slightly between the two years.

Section 4B In 1997-98 questions on seniority and hours worked are skipped for those in self-employed agriculture as there is a special set of questions for them in section 4F. An additional question is asked about whether the work is seasonal, and for wage workers, for how long the respondent has worked for this specific employer.

There are some differences in compensation items asked in the questionnaire.

1992-93 (Separately for cash and in-kind payments)

- a. Regular bonuses
- c. Lunch money (or in-kind)
- d. Tet and holiday bonuses
- f. Supplements for dangerous conditions, third shift, extra hours

Implicitly included in g.) other forms of payment

Implicitly included in g.) other forms of payment

Implicitly included in g.) other forms of payment

g. Value of other forms of payment for work

e. Subsidies for rice, electricity, bicycle

b. Bonuses for innovations, scientific research

1997-98 (Combined cash and in-kind payments)

- a. Regular bonuses
- b. Lunch money
- g. Tet and holiday bonuses
- c. Supplements for extra hours, extra shifts, dangerous conditions
- d. Supplement for on-the-job training
- e. Supplement for responsibility, seniority, frequent travel, talent, regional bonus
- f. Honoraria or royalties
- h. Other including (irregular bonuses such as innovations, subsidies for rice, electricity, bicycle, uniforms, blankets, warm clothes, relocation, per diem in Vietnam or abroad, etc.)

Explicitly prompted in h.) other

Implicitly included in h.) other forms of payment

In 1997-98 questions are also asked about salary payments in arrears, labor contracts and types of health benefits which are not asked in 1992-93. In 1992-93 a question is asked about payments from salary to maintain employment which is not asked about in 1997-98.

Section 4C In 1997-98 the question on total compensation payment for the secondary job prompts the interviewer to include the many types of compensation while in 1992-93 no such prompt is in the questionnaire.

Section 4F In 1997-98 a new section asks about weeks, days and hours worked in 6 different self-employed agricultural, agro-forestry, fishery activities. In 1992-93 hours worked was asked the same as for other forms of employment.

In 1997-98 No section exists on job seeking among the currently employed (past 7 days) as in section 4D of the 1992-93 questionnaire.

Differences with the 1992-93 Questionnaire
<p>Section 4G In 1997-98 new questions were added on age at beginning work, month and year last employed, whether the individual had any long periods off from work since they began working. Instead of asking about what the individual did prior to current employment, in 1997-98 a question was asked details about the job the respondent had 5 years previously.</p> <p>Section 4H In 1992-93 the questionnaire had questions about unemployment and job seeking in the past 12 months. In 1997-98 that section was replaced with a question about participation in corvee labor and various mass organizations.</p>

Section 5 MIGRATION

All household members age 10 and older responded to questions on migration in section 5. If not born at current place of residence, respondents were asked the province/country of their birth and whether the place of birth was a city, provincial town, district town, rural area or don't know. The age at which such individuals left their place of birth was recorded, as well as the main reason for leaving. In addition, individuals were asked the main reason for coming to the current place of residence, from what province/country and whether the previous place was a city, provincial town, district town or rural area. Respondents were asked the number of places they had lived for over 12 months, and if they had migrated for work in the past 12 months, for how many months and where they had gone.

Differences with the 1992-93 Questionnaire
<p>Province codes differ between 1992-93 and 1997-98.</p> <p>In 1997-98 instead of asking how many times an individual had migrated (as in 1992-93), a question was asked of how many places an individual had lived for over 12 months.</p> <p>The section on migration in the past 12 months is completely new.</p>

Section 6 HOUSING

6A TYPE OF DWELLING

6B HOUSING EXPENSES

6C HOUSING CHARACTERISTICS

Section 6 contains information on the type of dwelling, housing expenses, and housing characteristics for all households interviewed. Information was collected on the number of rooms in the dwelling, ownership status, rental cost if rented, drinking and washing water sources and distances to water, garbage disposal, toilet type, source of lighting, source of cooking fuel. Respondents for all households were asked for the resale value of the dwelling, regardless of whether the dwelling was owned or rented. This section also contains information on area of house, material for walls, floor, roof and doors and windows and current condition of the house.

Caution: In questions about electricity, some houses have small generators (hydro or gas) and therefore don't have electrical meters.

Differences with the 1992-93 Questionnaire
<p>Section 6B In 1997-98 questions were added on partial ownership of dwelling, co-owners of the dwelling, payments to co-owners.</p> <p>Codes for water sources changed distinguishing different types of wells and filtration. Questions were added on filtering water and boiling water.</p> <p>An additional code was added for toilets draining directly into the water.</p>

Differences with the 1992-93 Questionnaire (cont'd.)

Questions on distance and time to search for cooking fuel were dropped in 1997-98.

Additional codes for types of doors were added.

Section 7 RESPONDENTS FOR SECOND ROUND

In section 7, the principal respondent for Round One was asked to identify: 1) the household member who knows the most about all the agricultural and livestock activities of the household; 2) the four most important businesses and trades belonging to the household and the household members who know the most about them; 3) the household member who shops for food; 4) the household member who knows the most about the other household expenses, income and savings of household members; 5) a woman was selected at random from all women in the household between ages 15 and 49 who had ever been married to respond to the fertility module. See section 3.5 for more information on the sampling methodology used to select eligible women respondents.

In principle, those identified in this section for interviewing in later sections should be the ones who are actually interviewed in those sections. While this is true for many households there are some cases where the respondents for agriculture, food expenses, non-food expense sections are different from those identified in this section. This is possible if the person identified was not present at the time the section was completed (e.g. the second visit to the household).

Differences with the 1992-93 Questionnaire

In 1997-98 up to 4 non-farm household businesses could be selected compared to 3 in 1992-93.

In 1997-98 the woman selected to respond to the fertility section should have been ever married, while in 1992-93 any woman age 15-49 could be selected.

An additional question was asked about whether an individual selected to respond in the second round could not speak Vietnamese. The language that they speak was then recorded.

Section 8 FERTILITY AND CONTRACEPTION**8A FERTILITY HISTORY****8B FAMILY PLANNING**

In each household one ever-married woman 15-49 years old, randomly selected in section 7 responded to questions in section 8. If a household contained no woman in this age range who had ever been married, section 8 was not completed. The woman was asked if she had ever been pregnant, and if so, whether she had given birth. Women who respond that they have given birth are asked the birth date and sex of all children they have given birth to, including children who did not survive. If the child is not alive the woman is asked how long it survived. For all children born since 1988, the vaccination status is asked whether the child is still alive or not. The woman is asked about the birth of her last child, and for mothers of children age 4 and under questions on birth weight and breast-feeding are asked. All ever-pregnant women are then asked about abortions and miscarriages, whether currently pregnant and age at marriage.

Section 8B gathers information on knowledge, use, source and cost of seven modern and four traditional methods of family planning. In using data from this section it should be kept in mind that unlike the Demographic and Health Surveys and the World Fertility Surveys, interviewers are not necessarily women.

Differences with the 1992-93 Questionnaire

In 1997-98, only ever-married women were asked about fertility while in 1992-93 any woman age 15-49 was asked.

Section 8A In 1997-98 a table of vaccines is printed in the questionnaire to help identify which vaccines have been given. A longer list of vaccines is presented in 1997-98 and up to 8 different vaccines could be listed compared to only 4 in 1992-93.

Additional codes are added for family member and neighbor assisting in delivery.

An additional question is added on the length of delay between giving birth and breast-feeding and age at introducing foods other than mother's milk.

In 1997-98 a distinction is made between miscarriage and abortion.

Section 8B The list of contraceptive methods differs from 1992-93 dropping abstinence, insertion of medicine and menstrual regulation and adding use of spermicidal cream or foam with diaphragm and an "other" category.

Women not currently married are not asked about current use of contraception.

Additional codes for reason for never use of a particular contraceptive are added for religion and using other method.

Additional codes for reason for stopping use of a particular contraception are added for religion.

Second Interview

Section 9 AGRICULTURE, FORESTRY AND FISHERIES

9A1 ANNUAL CROP LAND

9A2 OTHER CULTIVATED LAND

9A3 LAND RENTED OR BORROWED IN

9A4 LAND RENTED OR LENT OUT

9A5 SWIDDEN LAND, ALLOCATION, EXCHANGE, INHERITANCE, PURCHASE, SALE, TRANSFER AND RETURN OF LAND

9B1 RICE

9B2 OTHER FOOD CROPS

9B3 ANNUAL INDUSTRIAL CROPS

9B4 PERENNIAL INDUSTRIAL CROPS

9B5 FRUIT CROPS

9B6 AGRO-FORESTRY CROPS

9C CROP BY-PRODUCTS

9D FARM EXPENDITURES

9D1 OTHER FARM PRODUCTION FACTORS

9E PROCESSING OF HOME-PRODUCED CROPS AT HOME

9F LIVESTOCK

9G OTHER ANIMAL PRODUCTS

9H AQUACULTURE

9I LIVESTOCK DEVELOPMENT SERVICES

9J LIVESTOCK EXPENDITURES (INCLUDING WATER PRODUCTS)

9K HANDTOOLS

9L FARM EQUIPMENT, MACHINERY, TOOLS

In section 9 the respondent was the household member identified in section 7 as the one most knowledgeable about the household's agricultural, forestry and fishery activities. Most questions refer to the past twelve months. This section is by far the longest section of the household questionnaire, with many subsections that contain information on different aspects of agricultural production and related livestock and fishery activities.

Section 9A1 to 9A5 collects information on household's control over different plots of land of different tenures. These include land allocated by the commune, auction land, privately held land, rented/sharecropped land, swidden land and land rented out by the household. Water surface, barren land, forestry land, newly cleared or reclaimed land, residential land, and roadside/riverside land are included. In each case data are obtained on total plot size, area irrigated, quality of land in the plot, number of years cultivated, type of land tenure, use of the land and fees for land use. In these sections data are also obtained on newly acquired land or land recently dispensed with including amount paid/received, area of land, year acquired or dispensed with, and type of land. Land quality levels with related tax rates are printed in the questionnaire to assist households to remember the quality of the land.

Sections 9B1 to 9B6 contain detailed output information for all crops grown by the household. This information is obtained separately for each crop and includes (in most cases) information on which plots of land and total area planted, quantity harvested including subcategories for amounts kept for seed, fed to livestock, paid to laborers or lost post-harvest. Amount of harvest relative to normal, place and amount sold or bartered are also included. In the case of paddy, information is obtained separately for the spring crop, autumn crop, winter crop, swidden rice, sticky rice and specialty rice.

Section 9C contains information on crop by-products including those consumed by the household and those sold.

Section 9D contains detailed information on use of production inputs including seeds, manure, chemical fertilizer, insecticides and herbicides for the 6 largest crops cultivated by the household. This information is crop-specific and can theoretically be linked with output information in the earlier sections by matching the data-sets by household and crop codes. Information on other inputs such as transportation, storage, rent of equipment and payment for services, labor exchange, hired labor and taxes are obtained at an aggregated level for each household. Labor days and amounts paid are separated by type of work.

Section 9D1 asks about inventories of crops not yet consumed or sold, and any agricultural extension services obtained by the household specific to each crop.

Section 9E contains information on transformation of home grown crops that were subsequently sold. This includes data on number of sales, revenues from these sales, and costs of production for each of 13 items. In some cases investments in equipment are included as costs of production in this section.

Section 9F collects information on livestock, poultry and other animals that are either consumed by a household or generate income. These data include an inventory of current number and value of animals possessed, number and value owned 12 months ago, number and value sold and purchased. The value of animals consumed, given away or lost and born or received as gift is also collected.

Section 9G collects information on value of animal products sold such as milk, eggs, silk, manure, etc.

Section 9H collects information on area of water surface being used for aqua-culture, value of animals sold and purchased in the past 12 months.

Section 9I collects information on extension services received in the past 12 months for animal husbandry or aqua-culture.

Section 9J collects information on costs for goods or services used in raising livestock or aqua-products and from whom they were obtained.

Section 9K collects information on hand tools, and section 9L collects information on implements and farm machinery owned by the household, and any rental revenues obtained from them.

Differences with the 1992-93 Questionnaire

Sections 9A1 to 9A5 are completely changed from 1992-93. Information is asked about each individual plot of land rather than for each ownership/crop type. In addition to basic information on land area and irrigated area and amounts paid for its use, questions are asked about quality of each plot, time allowed to use the land, and whether any services are included in the price to use the land. Land denoted as residential land is included in the 1997-98 questionnaire, but not in 1992-93. Detailed information on plots of land rented in and rented out by the households is collected.

Section 9B1 to 9B6 Questions are added on which specific plots of land, harvest as a percent of normal, and for rice, whether the amount sold is paddy or milled rice. Sticky rice and specialty rice are given their own line for information to be collected. There are more codes for where rice is sold. The question on amount paid to the cooperative is moved to section 9D. The question on reason for post-harvest loss is dropped. The codes for crops are different between 1992-93 and 1997-98.

Section 9C Two by-products were dropped: corn stalks and rice husks.

Section 9D Questions were added on hybrid seeds and percent of chemical fertilizer and insecticides purchased on credit. The questions on whether or not the crop using each of these inputs has been harvested was dropped. Codes for where inputs were purchased or credit obtained have been expanded. The crop-specific transport cost question was replaced with a question on total transport cost. More detailed information is collected on labor hired by households. Tax information is collected in the expenditure section rather than in the land section as in 1992-93.

Questions on sharecropping are now asked in the land rental section rather than in the expenditure section as there were few such cases.

Section 9E Tobacco drying was dropped. The question on which household members participated in agro-processing was dropped.

Section 9F Frogs and turtles were added in the snake category. Questions were asked not only about animals raised by the household, but also animals owned by the household. Questions were added on the number and value of animals held 12 months ago. For animals born, received as gifts, lost, dead, or consumed by the household, instead of asking about number of animals, the value of animals was asked.

Section 9J Payment for someone to watch the animals is now collected in section 9D. Payments for small repairs to animal husbandry facilities, insemination services were added.

Differences with the 1992-93 Questionnaire

Section 9L Durable equipment for storage was added as a category.
--

Section 10 SELF-EMPLOYED NON-FARM EMPLOYMENT

10A INFORMATION

10B EXPENDITURES

10C REVENUES

10D BUSINESS ASSETS

Section 10 gathers data on household businesses for up to the 4 most important enterprises operated by the household. The respondent for each enterprise is the household member most familiar with its operation (as identified in section 7). Data are gathered on the ownership, number of employees, type of employee compensation for each enterprise. For each business, expenditures over the last twelve months on wages, raw materials, and taxes are collected. The respondent is asked how much, in money and goods, was received from sales and how much of the enterprise's product was consumed by the household since the first interview. Information on ownership, sales and purchases of assets – buildings, land, vehicles, tools and other durable goods – in the last twelve months is also collected.

Differences with the 1992-93 Questionnaire

<p>Section 10A For any enterprise from four on up, total figures are requested of the household in 1997-98, but only the 3 most important enterprises were asked in 1992-93. Up to 5 household ID codes can be listed as working with any one enterprise. Questions were added in 1997-98 for the average wage for paid workers, whether the enterprise is registered, whether it represents another company.</p>
--

<p>Section 10B The question on wages was moved to section 10A. Packaging was added to the transport cost category and Insurance was added as a new cost category.</p>
--

<p>Section 10C Instead of just asking whether revenues in the past 2 weeks were more, less or the same as the previous 2 weeks, the respondent was asked for revenues in the past 2 weeks as a percent of revenues of the 2 previous weeks.</p>
--

<p>Section 10D Categories of assets were added including land, furniture. Equipment was separated from tools and vehicles was separated into large and small categories. A question was asked about whether each asset was used totally for production or sometimes allocated to household use, and what percent for personal use.</p>

Section 11 FOOD EXPENDITURES

11A HOLIDAY EXPENDITURES

11B REGULAR EXPENDITURES

In section 11A the amounts spent on holidays, primarily Tet (New Year), 15th January, 15th July, Moon Festival, and Independence day. The range of food items for which such expense information is obtained is smaller than that for which information is obtained in the section 11B. The main reason for separating holiday expenses from normal expenses, a departure from the standard LSMS survey format is to take into account the fact that the Tet holiday in Vietnam often represents significant departures from normal spending patterns – particularly unusually high expenditures.

Section 11B collects detailed information on market purchases and consumption from home production for 45 food items. Information is obtained for expenses since the interviewer's first

visit (two weeks previously). For a longer recall period (12 months) data are obtained on the number of months (in the past 12 months) each food item was purchased, the number of times purchases were made during those months, the quantity purchased, the number of times purchases were made during those months, the quantity purchased each time, and the value per purchase. These four pieces of information can be combined to obtain the total expenditure on food in the 12 months before the data of the interview. Note that this, in effect, is a variable-recall procedure because the time frame for which purchase information is provided by a respondent can differ for two food items, as well as across respondents. Besides market purchases (including barter), information is also collected on consumption from home production. Again data are obtained on the number of months each item was consumed, but unlike market purchases, the information of the quantity and value of consumption is obtained by asking a single question on the total amount for the last 12 months (as opposed to asking how often purchased each month, quantity purchased each time).

Differences with the 1992-93 Questionnaire

Section 11A Arrowroot noodles were replaced with specialty rice as a holiday food category. Meat items were converted to the weight without bones, internal organs or head. For the “other” category, the 1997-98 questionnaire prompts with examples such as processed foods, Tet cakes. The question about consumption of home-produced items in 1997-98 adds in non-purchased items (namely gifts)

Section 11B In 1997-98 the question on other fruits does not prompt for sapodilla or longans like in 1992-93. In 1997-98 the question prompts breakfast, lunch and dinner for meals eaten outside the home. As with holiday expenditures, in 1997-98 the question on consumption of home-produced items adds in non-purchased items.

Section 12 NON-FOOD EXPENDITURES AND DURABLE GOODS

12A DAILY EXPENDITURES

12B ANNUAL EXPENDITURES

12C DURABLE GOODS

12D EXPENSES FOR ASSISTANCE

Section 12 collects information on non-food household expenditures from the household member identified in section 7 as the one most able to answer non-food expenditure questions. In section 12A respondents were asked to recall the amount spent since the first interview (approximately two weeks) on daily expenses such as lottery tickets, cigarettes, soap, personal care products, cooking fuel, matches and candles, and gasoline. In section 12B expenditure data both in the last two weeks and the last twelve months were collected for shoes, cloth, clothing, home repairs, public transport, paper supplies, kitchen equipment, medical services, domestic servants, jewelry, entertainment and other goods (see household questionnaire). Purchase price, year of purchase and resale value of durable goods owned were collected in section 12C. Section 12D collects information on gifts the household makes to other households including the household member id, relationship to recipient, residence of recipient, and value of gift.

Differences with the 1992-93 Questionnaire

Section 12A Household level expenditures on tobacco products were replaced with individual level expenditures in section 3. Natural gas was added as a category. Dish-washing soap and other cleaning supplies were added to laundry soap category. An “other” category was also added. A question was added on home-production (and non-purchase, i.e. gifts received) of various daily non-food consumption items.

Section 12B A question was added on home-production (and non-purchase, i.e. gifts received) of annual non-food expenditure items. Spare parts, tires and maintenance for cars were added to those items for motorcycles. Paintings and house plants, purchase of gold, silver, precious gems for savings, cosmetic surgery and body building, corvee labor contributions, participation in revolving credit funds or purchase of stocks and bonds were added as a new categories. Health insurance was excluded from purchase of insurance as it is already asked in section 3. In 1997-98 the questionnaire was clarified so that wedding and funeral expenditures (code 160) are limited to those for household members while gifts to others for weddings and funerals are to be included in the gift category (code 162).

Section 12C Electronic games was dropped from the video player category. Record player was combined with the radio receiver. Computer, gas stove, fax, dehumidifier and vacuum cleaner, and electric generator were added as new categories. Pump was separated from hot water heater. Wall cupboards were separated from other cupboards.

Section 12D Relationship codes have changed to match those in section 1A. Questions were added on whether the recipient was male or female and for what purpose the gift would be used.

Section 13 REMITTANCES AND OTHER INCOME

13A INCOME FROM REMITTANCES

13B OTHER INCOME

Section 13 collects data on money and goods that come into the household as remittances or from other sources unrelated to employment such as social security, pension, poverty alleviation funds, interest on savings or investments, insurance payments, gifts, inheritance, lottery, renting out equipment or buildings, sale of vehicles or durable goods.

Differences with the 1992-93 Questionnaire

Section 13A A question was added on whether the sender of remittances was male or female and the use of the remitted funds. The relationship codes are different between 1992-93 and 1997-98.

Section 13B The category of social subsidies specifies clearly to include also the category of households who have war martyrs, war invalids, etc. A category is added for funds or programs related to the government poverty alleviation program and NGO, private or international assistance. Gold, silver and precious gems are added to the jewelry category, treasury bills and bonds are added to the other category.

Section 14 BORROWING, LENDING AND SAVINGS

14A BORROWING

14B LENDING

14C SAVINGS

Section 14A collects information on the amount of indebtedness of household members to people or institutions outside of the household. If money or goods have been borrowed, or borrowed and repaid by any household member in the last twelve months, information is collected on those loans, including the source and amount of loan, interest, side payments, collateral, repayment schedule, reason for borrowing, and number of loans from the same source. Section 14B collects similar information on the amount household members have lent to people outside of the household.

In section 14C, households are asked to list different types of savings, if any, including bank deposits, bonds, cash (dong), US dollars, gold and value of land. The respondent is also asked to total current value of all savings forms.

Differences with the 1992-93 Questionnaire

Section 14A In 1997-98 additional questions are asked including the code of the household member who borrowed money, if a revolving credit fund how many others participate, (and when it began), relationship, gender, province of residence of non-institutional creditors previous borrowing or lending from or to this creditor, what type of collateral was required, whether there was a guarantor for the loan and who it was, whether the loan has a fixed maturity, number of installments already paid back, separate questions on principal and interest already paid back, possibility of not being able to pay back the loan. In addition, the codes for relationship to creditor, purpose of loan, collateral have changed.

Section 14B In 1992-93 only two questions were asked about lending money out, whether anyone was indebted to the household and the total amount. In 1997-98 an entire section similar to the section on borrowing was added to the questionnaire.

Section 14C Respondents were asked whether they had ever utilized each of these forms of savings, while the 1992-93 questionnaire only asked about current use. The question on frequency of use of a form of savings was replaced in 1997-98 with a question on preferences for saving.

Several categories of savings were changed in 1997-98. The dollar category was clarified as US dollar to distinguish from other dollar currencies.

Savings in the form of paddy or rice was only collected from non-farm households who had not answered the questions in section 9D1 on inventories of agricultural products.

Land and housing information was only collected if it had not yet been included in section 6 Housing and section 9 agricultural land.

Several categories were added including: silver and other precious metals, foreign currency other than US dollar, precious gems, jewelry (including that not currently worn). Durable goods was dropped as it is already asked about in section 12C.

Section 15 ANTHROPOMETRICS

Anthropometric measurements are completed for each household member. Data were collected on the household member's age, gender, date of measurement, weight, height and arm circumference. It was also noted if female respondents were pregnant or breast-feeding. If a person was not measured the reason why is noted. There is no difference with the 1992-93 questionnaire.

2.2 Commune/Ward Questionnaire

A completely new commune questionnaire was developed for the 1997-98 VLSS survey with a greatly expanded content. A few questions in the 1992-93 questionnaire were dropped or moved to other questionnaires (see below). The commune questionnaire was administered by the team supervisor and completed with the help of village chiefs, teachers, government officials and health care workers. The questionnaire was administered only in rural and minor urban areas, i.e. communes 37 to 194, corresponding to villages 73 to 388 (see appendix G for complete list of survey communes and codes). Some sections of the questionnaire contain village/block level information, while most of the commune questionnaire refers to the commune.

The commune questionnaire contains 10 sections including:

0. SURVEY INFORMATION

1. BASIC PHYSICAL AND DEMOGRAPHIC CHARACTERISTICS

2. GENERAL ECONOMIC CONDITIONS AND AID PROGRAMS
3. NON-FARM SELF-EMPLOYMENT OPPORTUNITIES
4. AGRICULTURE
5. CREDIT AND SAVINGS
6. PHYSICAL INFRASTRUCTURE AND TRANSPORTATION
7. COMMUNE FISCAL INFORMATION
8. EDUCATION
9. HEALTH

Section 0 contains basic background information on respondents including position in the commune, education, gender and ethnicity. It also contains any comments about the interview or responses to different sections of the survey.

Section 1 collects information on population and geographic region of the commune. It also collects information on land area, population, number of households, ethnicity and religion of the population in the villages selected.

Section 2 contains information on major sources of income in the commune, reasons for changes in standard of living of the people of the commune, types of aid programs in the village, number of households considered poor by MOLISA standards, number of households receiving various types of assistance and general migration patterns.

Section 3 collects information on the existence of enterprises/factories, cottage industries or handicrafts operating in the commune. Information on basic wage rates paid by these various non-farm employment sources is also asked, separately for men and women.

Section 4 asks questions about the land distribution of the commune, main crops grown in the area, where produce is marketed and any problems affecting the harvest. Daily wage rates for various agricultural work for men, women and children under 15 were collected. The existence of agricultural extension center, various large farm machinery, and number of visits of extension workers were also asked about. This section also asked questions about natural disasters and environmental problems in the area.

Section 5 asks about opportunities for saving or borrowing for the commune.

Section 6 collects information at the village level about roads and waterways, and the distance and time from each village to various places such as the commune people's committee building, district capital, post-office, telephone, and market. Questions were also asked about any new infrastructure built in or near the commune, the main sources of water, whether there was a source of electricity in the commune and frequency of electricity outages.

Section 7 asked basic information on revenues and expenditures of the commune budget. Information was collected from the Commune accountant on various taxes, contributions and fees paid by each of the households surveyed in the rural clusters for the VLSS.

Section 8 collected information on distance and time to the nearest schools of each level from each of the village clusters. For each level of schooling, the reasons why children didn't attend school and the education problems of the commune were asked about. Information on anti-literacy programs was also collected. For each village, information on preschool was collected.

Section 9 gathered information on major problems with the health services in the commune and distance and time from the commune to each of various health facilities.

Contents of the 1992-93 Commune Questionnaire dropped in 1997-98

Note that most items from 1992-93 that were not included in 1997-98 have information available from other sources including the school, health center and household questionnaires)

- Number of women in the commune (total population compared with sex ratio from household information can substitute for this in 1997-98)
- Day and month of data collection of population figures (year is asked)
- Number of households in the commune (the 1997-98 questionnaire asked about the village level not the commune)
- Whether any households in the commune have pipe-borne water (information available in the household questionnaire in 1997-98)
- Whether there is a public loudspeaker or radio station in the commune
- What percent of households have a radio set or cassette player or TV (information available in the household questionnaire in 1997-98)
- Detailed questions on schools were moved to a school questionnaire
- Detailed questions on health were moved to a commune health center questionnaire
- Proportion of production given to the cooperative, government in total production harvested in the last 12 months (information available in the household questionnaire in 1997-98)
- For which items the people have to pay the cooperative (information available in the household questionnaire in 1997-98)
- Number of days of corvee labor required of households (information available in the household questionnaire in 1997-98)
- Whether most farmers in the community utilize chemical fertilizers (information available in the household questionnaire in 1997-98)

2.3 School Questionnaire

- I. GENERAL SITUATION
- II. FACULTY AND STAFF OF THE SCHOOL
- III. PHYSICAL FACILITIES
- IV. TEACHING EQUIPMENT
- V. EXAMINATION RESULTS
- VI. FINANCE AND SCHOOL FEES

The school questionnaires were implemented by the team supervisor to all schools within the two villages selected within a commune. There are between 1 and 7 school questionnaires filled in per commune. Section zero asks about the school name and code (as specified in section 8 of the commune questionnaire), whether the school is a satellite school or main school of a commune, what the main school code is if information was collected from the main school, the interview data and time and information about the respondent. Section one asks about the level of the school, ownership type of the school, number of shifts, periods, weeks taught in the year, whether lessons are taught in ethnic languages, number of classes, pupils, girl pupils, repeaters and for multi-grade classrooms how many pupils of each grade. Section two asks about teacher and director gender, qualifications and experience. Section three asks about classrooms and other amenities of the school. Section four asks about libraries and science laboratories. Section five asks about graduation examination results for each level of schooling. Section six gathers information about the various types of school fees for each grade and which if any pupils are eligible for reduction or exemptions from fees.

Differences with the 1992-93 Questionnaire

In 1992-93 a school questionnaire was not applied, but questions were asked in the commune questionnaire about schools. Questions from 1992-93 not asked in 1997-98 include for each education level: Number of students left at the end of the school year, whether textbooks are lent out for free or for a charge, whether there are sufficient textbooks for all students, number of pupils age 6-11, 12-14, 15-17 in the commune enrolled and not enrolled in the relevant level of education. Enrollment rates can be calculated from the information collected in the household survey and information at the commune level was therefore not asked regarding enrollments and children of schooling age.

2.4 Commune Health Station Questionnaire

- 0. SURVEY INFORMATION
- I. GENERAL SITUATION
- II. HEALTH WORKERS AND HOURS OF OPERATION
- III. EQUIPMENT AND SERVICES
- IV. ABILITY TO SUPPLY MEDICINES
- V. FEES

The commune health station questionnaire was implemented by the team supervisor. The respondent could be the director, doctor or physician's assistant of the health station.

Section zero asks about the interview date, time and position of respondent. Section one asks about the distance from the health center to the commune health station. Section two asks about staffing and hours of operation. Section three asks about in-patient beds, services provided by the health station and equipment available at the facility. Section four asks about availability of different types of medicines, and section five asks for the cost of various services, share of patients exempt from fees and person responsible for deciding on exemptions.

Differences with the 1992-93 Questionnaire

In 1992-93 a health station questionnaire was not applied, but several related questions were asked in the commune questionnaire. Questions from 1992-93 that are not included in the commune or health station questionnaire for 1997-98 include: Major health problems in the commune, whether there has been an immunization campaign, anti-malaria/dengue fever campaign, or health education for mothers in the past year, or five years. It is not clear why these were dropped.

2.5 Price Questionnaire

Price data were collected in all clusters, both urban and rural by the anthropometrist for 36 food items, 33 non-food items, 6 services, 10 pharmaceutical products, and 7 agricultural inputs. Three separate observations were made and these did not necessarily involve actual purchase. However, it is possible that as the anthropometrist is not a local person, the prices quoted are not the true prices of the locality. This information was utilized in checking unit prices in the consumption modules, and for calculating poverty lines. Price indices utilized for adjusting monetary figures to real values were obtained from the GSO CPI unit. Information on price indices can be found in Appendix D. Details on how and where prices were to be collected can be found in the anthropometry manual. The actual locations of price collection were recorded in the questionnaires, but unfortunately not entered in the computer files.

3 Sample

3.1 Sample Design

The target sample size selected for the 1997-98 Vietnam Living Standards Household Survey (VLSSII) was 6000 households. The majority of the sample was comprised of the households interviewed from the 150 communes selected in 1992-93 with the first VLSS survey (4800 households). Households are defined as people living and eating meals together in the same dwelling. In most cases there is only one household per dwelling as people who live together usually eat together. There were no clear indications in the various manuals on what to do if more than one household lived within one dwelling.

The sample in 1992-93 was a self-weighted sample drawn from all areas of Vietnam. The overall sampling frame was stratified into two groups urban and rural, with sampling carried out separately in each group (strata). According to the 1989 census, about 20% of Vietnamese households lived in urban areas so the sample stratification ensured that 20% of selected households also came from urban areas. Within each of the two strata, a list of communes was drawn up province by province from east to west and north to south. The selection of communes within each list was done to ensure that they were spread out evenly among all provinces in Vietnam.

Within each province in Vietnam, rural areas can be broken down into districts, districts broken down into communes and communes broken down into villages or hamlets. Urban areas consist of cities, provincial and small towns. Small towns are usually small areas and are divided into wards, and wards divided into blocks. Large cities and provincial towns are usually divided into districts, then wards and finally blocks. There were approximately 10,000 possible PSUs in 1992-93 (and 10,331 in 1998.) The average population in each PSU was approximately 6500. As will be explained in section 4, the workload for each survey team was to interview 32 households in 4 weeks.

The VLSS sample was drawn in three stages with communes/wards and small towns chosen as the primary sampling unit as that was the lowest administrative unit for which the GSO had estimates of population in 1992. A total of 150 communes/wards were selected systematically out of the 10,000 in all of Vietnam with probability of selection proportional to their population size. As some communes are quite large in size, logistically it would have been difficult to interview 32 households selected randomly within each commune/ward. Instead, population figures for each village/block were compiled from the selected communes to select two villages/blocks randomly with probability proportional to their population size. Finally, the third stage involved listing all households within each selected village/block and selecting 20 households (16 for the sample and 4 extras if it became necessary to replace a selected household). As there were two stages of sampling for which the sampling unit could be called a cluster, we have used the term commune/ward and village/block instead of cluster to designate the different units.

For the sample to reach 6000 households in 1998, an additional 1200 households were required. This was done by selecting households from the total sample of the 1995 Multi-Purpose Household Survey (MPHS) of the General Statistical Office (GSO). In order to provide an adequate sample to disaggregate results into the seven major regions for rural areas and three categories of urban domains, the selection of the additional households was not proportional to population, but instead was chosen so that the total sample of 6000 households over-sampled specific domains. In particular, the factors for over-sampling are presented in Table 3.1. It is essential in doing analysis to utilize the sample weights in the data to avoid biases in results due to oversampling of urban areas, and certain regions of the country. The weights are explained in section 3.3.

Table 3.1: Sample Allocation

Domain		Relative Sampling Fraction	Domain		Relative Sampling Fraction
URBAN			RURAL		
1	Ha noi, HCMC	2	4	Northern Mountains	1
2	Other cities	2	5	Red River Delta	1
3	Other urban areas	1.5	6	North Central	1
			7	Central Coast	1.5
			8	Central Highlands	3
			9	Southeast	2
			10	Mekong Delta	1

Because of the over-sampling factors and number of households in the existing VLSS sample, 96 households (three communes) had to be dropped from the Red River Delta in the VLSS share of the sample to come up with the approximate sampling allocation in Table 3.1.

Again, for survey implementation reasons, the households are not randomly selected throughout each domain, but are selected in groups of two villages/blocks in a commune/ward, and 16 households per village/block (15 households for the MPHS part of the sample). The final sample is therefore made up of 4704 households from the VLSS 1992-93 communes, and 1290 households from the MPHS survey communes for a total of 5994 households.

The household listings used for sample selection were provided by the local authorities, and in most cases consisted only of long-term registered households. As migration has increased in recent years, especially in urban areas and the Central Highlands, this could lead to some biases in the results as temporarily registered households would not have had a chance of being selected.

For a more detailed description of the sample design for VLSS 1997-98 see the paper on the recommended sample design by consultant Christopher Scott listed in Appendix C. For the list of sampled communes, see Appendix G.

3.2 Implementation

As explained above, the sample for the 1997-98 VLSS is based on households already selected and interviewed either in the 1992-93 VLSS or in the 1995 MPHS. However, in order to maintain the size of the sample, the 1997-98 VLSS instituted a procedure by which households in the original sample (1992-93 VLSS or 1995 MPHS) that were unavailable or refused to respond were replaced by another household selected randomly in their village/block. Procedures for choosing replacement households are explained in detail in the Supervisor's manual. Thus the total sample size is approximately the planned sample with exceptions explained below. The final sample allocation is given in table 3.2 by region and 3 categories of urban area.

Information was supposed to be collected in a supplementary small questionnaire which includes information on Section 0A, 0B and 1D, although in some cases this procedure was not followed. Among the households in the 1992-93 VLSS sample, 495 households were not re-interviewed in 1997-98. Ninety-six of the original households were dropped by not collecting information on three Red River Delta communes as mentioned above in sample design and no information is available on those households. For 46 other cases not re-interviewed, no information is available. For the remaining 329 cases some households had moved away (281 households), were temporarily away from the commune (19 households) were dissolved due to death or other reasons (1 household), refused to answer (12 households) or were not interviewed for some other reason (16 households) and had to be replaced. For these households, the survey team was asked to interview neighbors, relatives or commune authorities to understand the whereabouts of these households and their members and information is collected for section 0A, 0B and 1D of the

household questionnaire. Thus, of the 4800 original households from 1992-93, 4305 remain in the 1997-98 survey. For the MPHS sample we have information on 35 cases who were selected but not interviewed because the dwelling couldn't be found (9), were away from home (4), refused to participate (7) or other reason (15). It is possible that there were other households from the MPHS 1995 sample unable to participate in the survey, but no information is available for these cases.

Table 3.2: Sample allocation of households

Region	Large urban	Medium urban	Minor urban	Rural	Total
Northern Uplands	0	62	125	672	859
Red River Delta	188	126	78	783	1175
North Central Coast	0	92	16	600	708
South Central Coast	0	94	158	502	754
Central Highlands	0	0	0	368	368
Southeast	431	62	16	514	1023
Mekong Delta	0	62	220	830	1112
Total	619	498	613	4269	5999

In the case of one village in commune number 115, eight households remained in their original households, but the administrative boundaries had changed so the interview team replaced the eight households, but then went back and interviewed the original eight households, so instead of 5994 households, we in fact have 6002 households in the total sample. (This larger sample size in this cluster makes the probability of being sampled in that cluster 1 out of 40 rather than 1 out of 32. An adjustment has been made in the sampling weight calculation to take this into account). Because information is missing for large parts of the questionnaire for three households (1302, 7506, 19916) the final sample for the whole survey is adds up to 5999 (although information on some sections is still available for these 3 households). The original design has 190 communes. However, during the preparation of the survey, it was found that four communes had been split into two different administrative units, in three cases half urban and half rural, and in the third case a commune divided into 2 rural communes so codes were added for the new administrative areas (these are 1997-98 commune numbers 85 and 41, 111 and 45, 157 and 51, 43 and 44). Note for each of these communes there are only 15 to 16 households rather than 30 or 32. These additional codes bring the actual total number of communes up to 194, but each of these pairs of communes (8 in total) only consisted of 15-16 households. Ten communes from the MPHS were not separated into two separate villages but collected information on 30 households in the same village/block (communes numbered: 33, 49, 53, 141, 144, 158, 159, 162, 163, 185.). In a few clusters, primarily because of changes in borders of villages, the number of households in each village is not equal, notably several MPHS communes (codes 9,13,15,18,140) and commune 189 from the VLSS part of the sample.

3.3 Weights or Expansion Factors

In order to make estimates relating to groups of domains, or to the whole sample, the data *must be* weighted in order to correct the bias due to deliberate over- or under-sampling. This manipulation increases the sampling error slightly for nationwide estimates but allows us to do more precise analysis within each domain. The formula for calculating the weights (or expansion

factors) consists of the following factors to deal with the probability of the household being sampled at the different stages of sampling:

- The probability of being in the VLSS or MPHS part of the sample
- The probability of the commune/ward being selected out of the domain (strata)
- The probability of each of the two villages/blocks being selected out of the whole commune/ward.
- The probability of 16 (15 for MPHS) households being sampled out of the village/block.
- A factor to adjust for changes in number of households in the village/block between the original sample selection year and the time of the survey in 1997-98.
- Because there are two sample sources with different sample selection criteria, the formulas are slightly different for each component and are given below:

$$W_{dij}^{VLSS} = \frac{M_{d,VLSS}}{M_d} \cdot \frac{M_{0,d}}{c_d M_{0,di}} \cdot \frac{M_{1,di}}{2M_{1,dij}} \cdot \frac{M_{2,pij}}{16} \cdot \frac{M_{5,dij}}{M_{2,pij}} \quad (\text{Equation 1})$$

$$W_{dij}^{MPHS} = \frac{M_{d,MPHS}}{M_d} \cdot \frac{N_{0,d}}{c_d N_{0,di}} \cdot \frac{c'_d}{c''_d} \cdot \frac{N_{3,di}}{2N_{3,dij}} \cdot \frac{M_{4,dij}}{15} \cdot \frac{M_{5,dij}}{M_{4,dij}} \quad (\text{Equation 2})$$

With the following definitions:

Table 3.3

Symbol	Parameter
M	Number of Households
N	Population
<i>Subscripts</i>	
0	1989 ^a
1	1992
2	1992-93
3	1994
4	1995
5	1997-98
d	domain (strata)
i	commune/ward
j	village/block
<i>Other ^b</i>	
c _d	Number of communes selected in a domain
c' _d	Number of communes excluding overlap
c'' _d	Number of communes in sub-sample

^a The subscript 0 for the MPHS weight represents 1989 population figures updated to 1993

^b For the MPHS part of the sample, some communes overlapped with those in the VLSS sample, and were therefore not selected to add to the sample. This is taken into account by the variations of c_d.

The above formulas can be simplified by canceling terms. For the village (229) with eight additional households the denominator in the formula for that village should be changed from 16 to 24.

Therefore, any household observation contributes to the estimate of the total through the formula above. To estimate the total, we only have to multiply each observation by its appropriate weight and form the sum over all households in the complete sample. Once the weights have been computed and applied, there is no further need to distinguish the two samples. The weights for the VLSS 1997-98 can be found in **longlatfin.xls**.

3.4 Complications in calculating weights (expansion factors)

The process of calculating sampling weights (expansion factors) for the 1997-98 was very long and complicated due to missing information to plug into the sampling weight formulae, implausible results with given information and mistakes in the sampling weight formulae. The untimely death of Mr. Chris Scott, who designed the sample, complicated the final item because he was not available to consult about his formulae.

Missing Data problem

The missing data problem resulted from the fact that in both years of the survey, the commune questionnaire was not used for urban areas. Therefore information on the number of households and population in these clusters and residential blocks was missing and had to be obtained through contacting the provincial statistical offices which took some time.

The second problem was identified by implausibly large increases or decreases in the size of communes/wards and villages/residential blocks between 1992 and 1997 which led to extremely high or low weights. The high increases or decreases in numbers of households or population could be a result of higher fertility and in-migration or high mortality and out-migration. However, it is more likely that this is a result of mistakes in reporting information from the survey localities, or more likely, due to changes in administrative boundaries which make comparability a problem. The first solution to deal with these problems was to contact provincial statistical offices to obtain correct information or explanations of why there were large increases or decreases for localities reporting higher than 10% annual increases or decreases in the number of households in villages between 1992 and 1997. In many cases the provinces did respond, but in some cases they did not respond, perhaps because of the heavy work they are currently involved in to compile Census 1999 results. In cases where provinces responded, we used the new information, or in the case of Dak Lak, where the province said that there was a large in-migration, we accepted the numbers as given.

However, due to the fact that some provinces were unable to report the population of merged or split localities according to their 1992-93 boundaries, and because of the number of provinces not responding there continued to be a need to adjust some of the figures. In the cases where the ratio of households in 1997-98 to the number in 1992-93 in the villages was lower than 0.5 or higher than 1.5 and there was no clear explanation for a large population increase or decrease, we have made adjustments in one of two ways. For reference, these clusters are listed in the weights spreadsheet in bold which is included in the documentation available for the survey. The first round of adjustments was made by replacing the last item in the Chris Scott's weight equation with the ratio of households in the commune in 1997-98 over 1992-93. This was applied to all clusters that were in rural or small town areas for which information on the population of the commune was available. The number of households in the commune was estimated for 1997-98 by taking the population reported in the commune questionnaire and dividing by the average household size of the households selected for the survey in the sample cluster.

After applying this method, most of the ratios of households between the two years were within an acceptable range, and were used directly in the weight equation. However, 7 clusters still had ratios under 0.5 or over 1.5. In addition, for the 7 urban residential blocks that needed to be

adjusted, information on size of the commune was missing as the community questionnaire was not applied to them. In order to estimate the weight for these 14 clusters, annual population growth rate estimates were taken from Table 1.6, Variant 1, 1995 (i.e. growth from 1989 to 1995) in “Projections of Population, School Enrollment and Labor Force Vietnam, 1990-2005” This is a monograph from the Vietnam Population Census – 1989, published by the Statistical Publishing House in 1994. These growth rates by province were applied to an estimate of 1992 population taken by multiplying the number of households in the commune in 1992 by the average household size of the households in that commune (taken from household questionnaire 1992-93). The new commune population numbers for 1998 were then divided by the household size 1998 (taken from household questionnaire 1997-1998). The ratio of this estimated number of households in the commune over the 1992 figures were then used in place of the last item in the probability weight equation provided in Chris Scott’s revised sampling design report “Vietnam Living Standard Survey 1997-98,” written in July 1997. In one of these cases, the ratio slightly exceeds the 1.5 that we considered acceptable but we have left it as is.

For the MPHS part of the sample, we looked at the ratio of 1994 to 1993 population in the commune as the most likely to be wrong. We found indeed that there 12 cases with this ratio less than 0.5 or over 1.5. To adjust these, we took the annual population growth rate estimates from Projections of Population, School Enrollment and Labor Force Vietnam, 1990-2005. We applied to these to the 1993 commune population figures to estimate the 1994 commune populations, then plugged those estimated figures into the weight equations. The relevant communes are listed in bold in the spreadsheet.

Incorrect formula problem

In undertaking sample selection for the 1997-98 VLSS, 10 domains or strata were used and households were selected within each strata proportional to population. This and other sampling details are explained in great detail in Chris Scott’s report on revised sample design for the VLSS (1997). However, households were selected from two different samples, the VLSS and the MPHS. The problem is how to calculate the correct weights within any given domain. This includes three components, i) the probability that a commune was selected out of all the communes in that domain; ii) the probability that a household is selected given that it’s commune was selected; and iii) how to adjust the weights given that we are combining households from 2 separate surveys. Part ii) is quite straightforward and not problematic although there were some data problems as mentioned above.

To calculate the probability of a commune being selected, Chris Scott’s formula included the inverse of the ratio of the number of households in a given commune over the number of households in the province. However, given that sampling was done based on domains, rather than provinces this is not strictly correct. It is quite simple to fix this as we know that for the 1992-93 survey, communes were selected with probability proportional to number of households in the commune. We can then take the ratio of the number of households in the commune over the number of households in the given domain (of 10 domains) to get the probability of the commune being selected. For the MPHS portion of the sample, within 7 provincial size domains, provinces were selected with weights proportional to their population within those domains, then communes were selected in a way that within provinces weights were given to communes that are proportional to their populations within those provinces. Thus within any new domain system (10 domains for the VLSS II) for the MPHS households, the probability of a commune being selected is simply the population of the commune divided by the population of the domain.

To adjust the weights given that we are combining households from 2 separate surveys. For the three domains which have no households taken from the MPHS survey, the adjustment factor is just 1, as 100% of the sample in these three domains is from the VLSS. For the remaining 7

domains, we use an adjustment factor equal to the share of households that comes from the VLSS or the MPHS from each of these 7 domains.

In concrete terms, then, the following adjustments were made to the weight equations 1 and 2 above.

The provincial level terms were converted to domain (strata) level terms (i.e. households (M_{op} changes to M_{os}) in the domain for VLSS part and population (N_{op} changes to N_{os}) in domain for MPHS part).

The numbers of communes selected (C_p) are now the numbers of communes selected in the domain (strata) (C_s) for both the MPHS and VLSS.

The fraction of the sample 4800/6090 and 1290/6090 are changed to be the share of households in the specific domain that come from the VLSS or from the MPHS surveys and applied the relevant weights.

As information on the households in the 10 domains is only readily available for 1993, we have used these figures along with numbers of households in the communes for 1992 to estimate the probability of a commune being sampled rather than the 1989 figures. While this is not strictly correct, as the size of domains is growing at different rates, it is approximately correct and should not have a large impact on the weights.

Evaluation of Weights (Expansion factors)

There are several ways to check that the weights make sense as shown in Table 3.4.

First, the sum of the weights over all the households should equal the number of households currently in Vietnam. The sum of the weights times household size should approximately equal the population of the country. The number of households in 1993 was approximately 14.5 million households. The sum of weights gives us approximately 16.1 million households. With population growth and a slight decline in average household size over the 5 year interval between surveys from 4.97 to 4.75 we would expect an increase of about this amount. If we estimate the total population from the sample weights times household size, we come up with approximately 75.65 million people. This is approximately the number of people between 1996 and 1997. However, the 1999 preliminary Census results show that population is lower than projected population by about 1 million people so this result is quite reasonable.

The second criteria for assessing the weights is that the proportion of the weights coming from each of the 10 domains should be equal to the proportion of households in Vietnam coming from each of those domains. The structure in column A is taken from Table 1 of Chris Scott's paper and is based on 1993 estimates of population. The shares we get from our weights are shown in column B. We can see that while there are some differences in the structure, they appear to be reasonable given migration patterns in Vietnam. Namely, we see in-migration to large urban areas (domain 1), very small urban areas (domain 3) and the rural Central Highlands (domain 8), and out-migration from other regions, especially the Mekong Delta (domain 10). This migration would affect the last factor in the weight formula which is the ratio of the number of households in the village in 1997-98 over the number of households in the village in 1992-93 for the VLSS part of the sample, and over the number of households in the village in 1995 for the MPHS part. The migration patterns mentioned above should lead to the share of households in these in-migration regions being higher than their 1993 shares.

Table 3.4: Criteria for Evaluating the reasonableness of Weights for VLSS 1997-98

Domain	Share of households 1993 (Table 1, Scott)	Share of households 1997-98 from weights	VLSS average weight	MPHS average weight	Ratio of mean weight to Red River Delta	Original sampling fractions in C. Scott
	A	B	C	D	E	F
1	6.20	7.55	1896	2479	1.6	2
2	5.51	5.08	1744	1502	2.1	2
3	8.21	8.99	2579	3041	1.3	1.5
4	14.72	14.52	3477		1.0	1
5	19.11	18.76	3494		1.0	1
6	12.60	12.44	3428		1.0	1
7	7.69	7.62	2190	2571	1.5	1.51
8	2.70	3.06	1232	1392	2.6	3
9	6.00	6.04	1746	1721	2.0	2
10	17.26	15.95	3089	3196	1.1	1

The third criteria for evaluating the weights is that within domains, the mean weights given to MPHS and VLSS households should be approximately the same. We find that the weights are close, but only for domains 9 and 10 (Southeast and Mekong Delta) do we find no significant difference in the means of the weights. However, we don't find any real pattern such as MPHS weights always being higher or lower than VLSS weights. In domains 1, 3, 7, 8, 10 (Large cities, small towns, South Central Coast, Central Highlands and Mekong Delta) we find that MPHS weights are higher than VLSS weights, while in regions 2 and 9, MPHS weights are lower than VLSS weights. Results based on this criteria are not particularly strong in favor of the weights that have been calculated. However, if the other three criteria indicate the weights are legitimate, we can probably accept the weights.

The fourth criteria for evaluating the weights is that within domains, the mean weights should be proportional to the over-sampling weights given for each domain in Table 3 of Chris Scott's paper. Differences from these over-sampling fractions should be due to migration and differential fertility. For example, the mean weight for households in the Red River Delta with a sampling fraction of 1 should be 3 times the mean weight in the Central Highlands. Deviations for specific communes or villages should also be attributable to differential growth since 1989 in those communes or villages.

Sources of information for calculating weights

Table 3.4: The sources of information to calculate weights were found from various sources:

Symbol	Data needed	Source
$M_{o,d}$	Number of households in the domain in 1989	Census results; As 1989 results are not compiled for the 10 domains, we used estimates of numbers of households in 1993 which were reported in Chris Scott's report.
$M_{o,di}$	Number of households in the commune in 1989	Special Tabulations of Census results by commune (1993 figures were used to be comparable to the figures for the total domain)
$M_{1,d}$	Number of households in the commune in 1992	Commune questionnaires for rural areas; estimates of population size of communes

		made from 1989 Population census results, divided by average household size in the cluster taken from household questionnaire results.
$M_{1,dij}$	Number of households in the village in 1992	Special administrative records kept by GSO from the 1992-93 VLSS
$M_{5,dij}$	Number of households in the village in 1998	Commune questionnaire for rural areas and small towns; Special request to the provinces for this information for urban areas.
$N_{0,d}$	Population of domains using estimates from 1993 based on 1989 Census	Chris Scott's report, 1997
$N_{0,di}$	Population of commune using estimates from 1993 based on 1989 Census	Tabulation kept by the Social and Environmental Statistics Department for sampling of the MPHS
$N_{3,di}$	Commune population 1994	MPHS administrative records for sampling
$N_{3,dij}$	Village population 1994	Estimated by assuming that the ratio of village population to commune population in 1994 is the same as it is in 1997-98. Rural village and commune population 1997 taken from commune questionnaire. Urban commune population in 1997 requested from provinces.
$M_{5,dij}$	Number of households in village in 1997	Commune questionnaire for rural areas and small towns. Special request to the provinces for this information for urban areas.

Note: A few items canceled each other out and are therefore not needed in making weight calculations.

As it is difficult to obtain 1989 figures for numbers of households in the 10 domains in 1989, the ratio of number of households in the domains over the number of households in the commune from 1993 were used instead. It is possible that this ratio changed somewhat between 1989 and 1993, but not significantly. If we can obtain information for the numbers of households in the domains in 1989 we can change this.

3.5 Selection of respondent for the Section on Fertility

The survey collected fertility information for one woman per household. The woman, aged 15 to 49 who had ever been married, was selected at random from among household members, i.e. each eligible woman in a household had the same probability of being selected. The method used was to generate random permutations of household members' personal identification numbers on adhesive labels. Each questionnaire was affixed with a label in section 7. In order to select a woman at random, the interviewer scanned the list of personal id codes on each label until arriving at the id number of the first eligible female household member. Besides providing a random means of selection, this methodology ensured that the interviewer's selection of the respondent was verifiable and replicable by the supervisor.

4 Organization of the Survey

4.1 Survey Management

Decision 636 of the GSO director on 12 December 1997 serves as the legal basis for undertaking the VLSS. Under this decision, a steering committee consisting of 8 people was established. Mr. Nguyen Van Tien, Deputy General Director of the GSO was assigned to be the head of the

committee as he is in charge of social statistics at the GSO. Mr. Nguyen Phong, Deputy Head of the Social and Environmental Statistics Department was the standing director for the steering committee. The Director and the other deputy director of the Social and Environmental Statistics Department were also members of the steering committee. The GSO Chief of the Secretariat was on the steering committee and played an important role in logistics. The head of the Personnel and Training Department was involved in mobilizing the interviewers and supervisors at the provincial level. The head of the Methodology department was involved to supervise the methods used. The head of National accounts was in the steering committee to ensure consistency with national accounts as the VLSS data are heavily used in calculating the national accounts.

At the regional level, senior staff of the Social and Environmental Statistics Department were involved in direct supervision and backstopping for the field teams. The provinces were requested to support and cooperate with the survey in all aspects. In this regard, the provincial statistical offices played an important role in contacting the authorities in each locality to explain the purpose of the survey and to solicit cooperation from the local authorities.

The survey itself was carried out by 19 teams, each composed of 5 members: one supervisor, two interviewers, one anthropometrist, and one data entry operator. Each team worked within a set of provinces, and the teams were spread out over the whole country for most of the survey so the data at any quarter could be considered equally distributed across the country, and in urban/rural areas. The exception is the first cluster which was surveyed prior to Tet, where only four teams worked in the southern regions, with 15 teams in the north. This unequal regional distribution of teams was due to the fact that most interviewers were from Northern provinces, and the cost would be too high to transport them back and forth for the Tet holiday.

As in other LSMS (World Bank) surveys, the data entry was done in a decentralized manner (described below), which allowed for interviewers to return to households to check errors detected by the data entry program.

4.2 Training and Field Testing

The draft questionnaire was first field tested in the summer of 1996 in Thanh Hoa province in the northern part of Vietnam, but then the project was halted as the Government questioned the need for such a survey when the GSO already undertook an annual socio-economic survey of households (Multi-Purpose Household Survey). The project was redesigned so the VLSS served not only as a source of socio-economic data comparable to the earlier VLSS in 1992-93, but also as a training tool to build capacity along with significant training activities. After one year from the first pilot test the project was approved and in July 1997 a second pilot test was implemented in HCMC. In the fall of 1997, Professor Loren Brandt from the University of Toronto visited the GSO and provided useful assistance in revising the agricultural, agricultural labor and credit sections of the questionnaire after several field trips in different regions of the country. The pilot tests held in both the north and south allowed regional differences to be better taken into account in the survey. A four-week training course for interviewers, supervisors and data entry operators was held in Hanoi from 3 November to 3 December. The training included actual household interviews in one urban and one rural area for all trainees. Tests were given at the end to select the best interviewers, anthropometrists, data entry operators and supervisors. These and other local activities were done with the assistance of a long-term coordinator in Hanoi (Lisa Drummond (1996) and Sarah Bales (1997-1999). Guidance and short-term visits were undertaken by Paul Glewwe (World Bank).

4.3 Organization of the Field Work

Field work

Before any survey teams arrived in survey clusters, the head of the Provincial Statistical Office of each province sent staff to make contact with the commune/ward authorities to announce and explain the purpose of the survey, as well as to solicit their cooperation. When the survey team arrived in a commune (rural areas) or ward (urban areas), the supervisor met with local authorities to introduce the team, and if necessary explain more details about the survey. The supervisor then consulted with the commune authorities and village/block cadres to determine which of the original VLSS (or MPHS) households remained in the area as distinct households. Lists of all households in each village/block were compiled to provide a basis for the supervisor to select replacement households when necessary. Households were asked to come together for a meeting to introduce the survey team, the purpose of the survey and to explain what would be required of the households.

The household questionnaire was completed by the interviewers in two separate interviews. The first visit (Round 1) covered sections 0-8, while the second visit covered sections 9-14. Section 15 (anthropometric measurements) was completed by the anthropometrist during both visits. The typical interview time for completing half of the questionnaire (one round) was intended to be 2-3 hours. In fact it took longer and the mean total interview time was 7.4 hours, 7.5 in rural areas and 7 hours in urban areas in two separate visits for one completed household questionnaire.

Each team completed 32 household interviews during each four week period, 16 each per team interviewer. Round 1 (sections 0-8) was conducted in one village/block in the first week. The next week consisted of completing Round 1 in the other village/block in the same commune. In the third week the team returned to the first village/block and completed the remainder of the questionnaire (sections 9-14) for each household and the same was done in the second village/block in the fourth week.

The decentralized data entry was done as follows. After completing the first half of the household questionnaire (Sections 0-8) in the first village/block during the first week, all 16 half-completed questionnaires were given to the data entry operator (DOE) located at the team's "base". The DOE entered the data on a personal computer using the customized data entry program. The data entry program performed range and consistency checks on all data in the questionnaire and produced reports (lists) of all data for which problems arose. Upon returning to the same village in the third week to complete the questionnaire (and similarly upon returning to the second village in the fourth week) discrepancies detected by the data entry program could be corrected. After four weeks of work are completed in a given commune, the teams move to the next commune.

Duties of the supervisor also included observing some interviews and randomly revisiting some of the interviewed households to check the quality of the work done by the interviewer. A complete description of these and other supervisory tasks are found in the Supervisor's Manual (See Appendix C)

In addition to supervisory visits by members of the steering and standing committees, two consultants (Sarah Bales and Lisa Drummond) made supervisory visits to all teams in the first two months of the survey. Some additional reminders and guidance were sent out to all teams after these visits and this document is available from the World Bank (see Appendix C of this document.) During these visits some interviewers and supervisors were dismissed as they were not able to keep up the pace or maintain the quality of the survey work.

Timing

The field work for the VLSS began on 24 December, 1997 and was completed on 19 December, 1998. The quarter listed in the date of first interview and region of residence of the household are listed in Table 4.1. While an effort was made to distribute the clusters throughout the country during each month of the survey, it was difficult to ensure this because of the high transport costs involved. Each team was assigned 10 clusters. The order of surveying these clusters was determined so as to minimize the transport costs between clusters. The 19 teams were spread throughout the country during the whole period of the survey. However, in the first cluster before Tet, fewer southern clusters were done so as to avoid having to transport teams from the north down to the south and back after one month. In the Mountainous regions, care needed to be taken to avoid the rainy season when transport becomes next to impossible.

Table 4.1: Quarter of First Interviews by Region of Residence

Region	Quarter of Data collection				Total
	1	2	3	4	
Northern Uplands	29.10	25.03	38.42	7.45	100.00
Red River Delta	39.91	19.32	7.74	33.02	100.00
North Central Coast	21.47	24.44	18.79	35.31	100.00
Central Coast	21.62	19.89	28.78	29.71	100.00
Central Highlands	11.68	46.20	33.42	8.70	100.00
Southeast	21.99	28.25	25.32	24.44	100.00
Mekong Delta	17.90	25.63	30.94	25.54	100.00
Total	25.02	25.15	24.95	24.87	100.00

The VLSS dataset is collected with many reference periods of one year in order to avoid seasonal biases of asking about the past 4 weeks. However it is nevertheless likely that people will recall based on recent events. Therefore, for analysis, it might be necessary to know when important events occur such as crop planting and harvesting. In the 1997-98 dataset, question 7 in section 4 of the commune questionnaire asked about the month of planting and harvesting for main crops. During the time of the survey, one Tet holiday was celebrated on 28 January 1998 and it is likely that expenditures around Tet were higher than normal.

Note of caution: In December 1997, there was a severe, widespread typhoon in the Mekong Delta region, an area which sees frequent flooding, but seldom sees typhoons. While results for the survey are in most cases for a 12 month reference period which should help to even out the impact of this natural disaster, it is possible that it nevertheless affected survey results for this region, especially for interviews that occurred early in 1998.

5 Using the Data

It is strongly recommended that the data be used with the questionnaires. The questionnaires contain the exact wording of the questions and interviewer instructions. Since the questionnaires used to collect the data were in Vietnamese, it is also advisable to check the Vietnamese version of the questionnaire in case of doubt or confusing regarding a question. Sometimes the list of responses was to be read to the respondent, but more often the interviewer was simply to code the response given. The questionnaire is also useful in interpreting the codes. All codes are contained in the questionnaire itself. For more details on coding see the interviewer manual.

The most important reason to consult the questionnaire is that extensive use is made of skip patterns. This was desirable to maximize the ease with which the interview could be conducted and to include all questions that applied to a particular household or individual but exclude those that were irrelevant to a particular respondent or household. The researcher must be aware of these skip patterns so that the data are properly interpreted. The skip patterns are in most cases clear. If there is no instruction the next question should be asked regardless of the response. An

arrow followed by a number in parentheses (e.g. (>>21)) after a particular response indicates which questions should be asked if that reply is given. This implies skipping over other questions. An arrow with a number in a rectangle below the whole list of responses indicates which question to ask next regardless of the response. In addition, in each section there are implicit skip patterns which lead to missing information due to non-applicability. These include age restrictions on respondents for each section, exclusion of non-household members from all but section 1A, and non-consumption or non-ownership of particular consumption or asset items. Copies of all questionnaires (English and Vietnamese), and the manuals (interviewer, supervisor, data entry operator, anthropometrist, commune, school and clinic questionnaires) are available from the World Bank (See Appendix C).

The household, community, school, clinic and price data are available, to those who have received permission for use, in data-sets that correspond to sections of the questionnaire. A complete list of data sets is provided in Appendix B. The data sets are available in STATA (version 5.0). The STATA files contain variable labels for most variables in Vietnamese. SPSS and SAS format datasets are being created.

5.1 Data Processing

This section describes the different data processing stages between the actual interviewing and the final data-sets that are ready for use by researchers. The main reason for being aware of these data processing steps is that in case of questions about data quality, the first place to look for answers should be the process through which data passed from the questionnaire to the final data set stage.

The Vietnam Living Standards Survey format calls for two rounds of interviews and data entry in the field itself. The purpose of this format is to minimize response and data entry errors, and offer the possibility of rechecking information with households. Achieving these objectives requires continuous monitoring on the part of the supervisors of each survey team. The quality of interviewing and supervision varied from one survey team to another, but on the whole the survey effort was very good.

The procedure that the supervisors are supposed to follow is the following. Data collected in the first round is checked by the supervisor prior to the second round and necessary clarifications sought from the concerned household during the second round. Since the bulk of expenditure data are collected in the second round this particular checking is not possible for expenditure information. But since data are entered in the field office itself, the supervisor can check data from the second round, after the interview, and, in theory, check with households for which there are questionable data.

All responses obtained from individual, household, and community level interviews were recorded in questionnaires. In cases where the respondent didn't know the interviewers recorded DK in the questionnaires. Data entry operators were trained to input this as a period (.) which represents missing information in the dataset. No other special codes were used for "no response" or "don't know". All the data were entered into the computer in the field using data entry programs written in BASIC. The programs for the household survey were written by the GSO staff jointly with an international database consultant. The programs for the community questionnaire were written by a Vietnamese consultant with experience from the 1992-93 VLSS. The data produced by the data entry program are in the form of household files, i.e. one data file for all of the data in one questionnaire (whether household, community, school, or clinic.) In the case of the household and price surveys implemented in all communes/wards, this means 6002 data files and 190 files (according to the original codes for communes) respectively. The commune, school and clinic surveys were only implemented in rural areas, but more than one school was interviewed in most clusters. There were therefore 156 commune files, 155 clinic files

and 574 school files. These are the files that were produced in the field and then sent to the GSO in Hanoi.

At the GSO in Hanoi, these data files were processed to produce STATA data-sets, each of which contained information for all households for a subset of variables. The subset of variables chosen corresponded to data entry screens, so these files are hereafter referred to as “screen files”. For the household survey component, 115 data files were created, for the community survey component, 59 data files were created consisting of 37 in the commune questionnaire, 12 in the school questionnaire, 5 in the clinic questionnaire and 5 in the price questionnaire. The project coordinator trained members of the Social and Environmental Statistics Department in the use of STATA to check, correct and analyze survey data. The survey team then intensively checked the data for 5 months, entering data that had been missed and correcting data by checking the questionnaires for original recorded information. Data cleaning continued until the end of 1999 as the staff of the GSO produced tables for the statistical abstract, calculated the poverty line and participated in analyzing the data for topical reports.

In the training it was emphasized that correction in the raw data set should only be made in the case of incorrect data entry. However, there were other mistakes where codes had been entered incorrectly, but there was evidence within the questionnaire itself for repairing the mistakes, such as gender coded differently by the interviewer in different sections, but easily identified by the name of the individual, or the id code of a man recorded in section 8 when only women were eligible for answering that section. In these cases, STATA programs were written to correct these mistakes in the final data set available to users. These programs can be obtained upon request from the GSO. We have no reason to believe that any corrections were made in the data in order to obtain a politically motivated result.

The data have been used by the GSO to produce a statistical abstract, and a set of topical reports. Together with the World Bank, the GSO has calculated poverty lines for Vietnam which have been used in recent reports on poverty alleviation in Vietnam. No serious problems with the data have been uncovered. Some issues related to expenditure data are explained in Appendix E.

5.2 Linking Parts of the Household Questionnaire

Each household has a unique identification number which is contained in the variable *househol*. Values for this variable range from 101 to 38820. The first 1 to 3 digits are the code for the village/block. It is easy to derive the code for the village/block or commune/ward from the household code. In STATA one could use the following formulas:

```
gen village=int(househol/100)  
gen commune=int((village+1)/2) or  
gen commune=int((int(househol/100)+1)/2)
```

The last two digits of *househol* contain the household number within the village/block. For most clusters, these range from 1 to 20. For 10 MPHS clusters, 30 households were selected from only one village so the household numbers extend to at least 30, but can extend as high as 37. Commune and village numbers are completely different from the 1992-93 codes, but an attempt was made to utilize the same 2 digit code for the households in the panel. Unfortunately, in a few clusters the interview teams did not understand this and recoded the households completely. This is not a problem because 1997-98 survey households can still be linked to the 1992-93 households with section 0B.

Data for households from different parts of the survey can be merged by using the *househol* variable which appears in each data set of the household survey. To link information for an

individual one must use both the household identification number, *househol*, and the person identification number, *idcode*.

A child in the household can be linked to the parents, if the parents are household members, through the parents' id codes in section 01B. For parents who are not in the household, information is collected on the parent's schooling, main occupation and whether he/she is currently alive. Household members can be linked with their non-resident children through the parents' id codes in section 01C. The woman chosen for the fertility section can also be linked to the children living with her in the household through the id codes in section 8A. There is no code to link the children in her fertility roster who are not living in the household with the non-resident children in section 01C, though almost all can be linked by comparing age, sex and mother's *idcode*. In the same way spouses can be linked to each other through the spouse's id codes available in the household roster.

For panel analysis, household codes in 1992-93 can be derived from information in question s0bq02 and s0bq021 in section 0B of the 1997-98 household questionnaire. The 1992-93 code can be derived with the following formula in STATA:

gen HID92=s0bq02*100+s0bq021

Information to link individuals in 1992-93 and 1997-98 can be found in section 1D. *vlsscode* is the variable for the 1992-93 individual identification code and *idcode* is the corresponding 1997-98 individual identification code.

5.3 Linking the Household to the Price and Community Data

All community questionnaire pages have commune and some have village (cluster) identification codes which can be used to match commune or village level information with household information. The variables commune and cluster within the household and commune data files refer to the commune and village codes respectively. In the data files cmt072 and cmt073 on household level contributions to the commune, the variable *hhc* refers to the number of the household within the cluster. Thus, for village 75, hhc number 1, the corresponding *househol* would be 7501.

When merging files, it is always a good idea to check the range of values on the different variables prior to merging. In the household level data, village (cluster) codes for range from 1 to 388 and commune codes range from 1 to 194. Village codes in commune, school and clinic questionnaires begin at 73 and commune codes at 37 as they refer only to small towns and rural areas. See section 5.2 for the formulas for deriving the commune and village codes from the household identification number.

Note of caution: Because of changes in the coding of some administrative areas that were split, no commune level information is listed for commune 41 and 45. However, these communes/wards correspond to the original commune/ward numbered 85 and 111 respectively. For the price questionnaire, information is missing for communes 41, 44, 45 and 51. However, these communes correspond to current codes 85, 43, 111 and 157 respectively as they are the half of the commune that was split off.

6 Constructed Data Sets

Staff at the General Statistical Department with assistance from the World Bank have created a data set with information on annual household expenditures, region codes, etc. This constructed data set is made available for general use with the understanding that the description below is the only documentation that will be provided. Any manipulation of the data requires assumptions to be made and, as much as possible, those assumptions are explained below. Except where noted,

the data sets have been created using only the original (raw) data sets. A researcher could construct similar data-sets incorporating different assumptions.

6.1 Aggregate Expenditure

The data set *hhexp98n* contains variables for household annual nominal expenditures, household annual real expenditures, regional and monthly price indices, per capita nominal and real expenditures, and variables for the different components of total household expenditures – food expenditures, non-food expenditures, use value of consumer durables, etc. These along with the items used to calculate household expenditures are detailed in Appendix E. The data set also contains variables for a household's commune and village number which can be used to match household data-sets with community and price data-sets. In addition to this commune number, the data set also contains variables for regions, rural/urban, and province.

6.2 GIS Information

Information on latitude, longitude and altitude of each commune/ward selected in the sample of the VLSS in 1997-98 was compiled by the Office of Geology and Minerals. This information allows the VLSS to be linked to geographic specific information from other sources. File *longlatfin.xls* has the longitude, latitude and altitude coordinates of all clusters in the 1992 and 1997 VLSS surveys with the cluster names. *Longlat.dta* contains the coordinates and the commune codes for merging with other data to do analysis in STATA.

Appendix A. How to Obtain the VLSS 1997-98 Data set

Permission to utilize the VLSS

The VLSS data are the property of the government of Vietnam. Permission to use the VLSS data must be obtained from the General Statistical Office in Hanoi, Vietnam. The request should be submitted to:

Mr. Nguyen Phong
Department of Social and Environment Statistics
General Statistical Office
2, Hoang Van Thu Street
Hanoi, Vietnam
fax: 84-4-846-3511 or 84-4-846-4345
e-mail: vie95043@undp.org.vn

The request should include a brief (one page) explanation of the proposed research.

Tentative Fees for use of the data

Fee for one data set

A. For Vietnamese individuals and Vietnamese organizations

- Vietnamese organizations with domestic budget: \$US200*
- Vietnamese organizations with financial support from foreign and international budget: \$US400*
- Vietnamese individual with own pocket money: \$US100

B. For foreigners/ foreign or international organizations:

- Foreign or international organizations: \$US2000*
- Individual foreigners using the data set for own research/study/lectures preparation:
 - + Citizens of a developing country: \$US200
 - + Citizens of a developed country: \$US500

Fee for two data sets:

1.5 times higher than the fee for one data set

*In the case of institutions, the fee paid allows many people/projects belonging to the institutions to utilize the data.

Payment

An additional \$US20 fee is required for shipping the CD-ROM by airmail.

The procedure on getting the VNLSS data sets is as follows:

Pay the fee for data set(s) to the following account with a bank transfer. **DO NOT SEND PERSONAL CHECKS:**

Account name: TONG CUC THONG KE

Account number: 001 1 37 008159 0

Bank name: BANK FOR FOREIGN TRADE OF VIET NAM

Bank address: 23 Phan Chu Trinh Street, Ha Noi, Viet Nam

Once the bank confirms that the payment for the fee is available in the account, a CD-ROM with data set(s) and related materials will be sent to you by airmail.

Conditions of use

The user must respect the copyright of the data and should not allow other users to use the data without permission of the GSO. The user is requested to send to the GSO the results of use of the data at least one month before their publication.

Appendix B. List of VLSS Data Sets Available

The following data-sets are available on CD ROM in STATA version 5.0.

Questionnaire	VLSS DATASETS [*.dta]
HOUSEHOLD (numbers correspond to sections of the household questionnaire) [6002 obs.]	scr00a,scr00b,scr00c,scr00d
	scr01a1, scr01a2, scr01b, scr01c1, scr01c2, scr01d
	scr02a, scr02b, scr02c, scr02d, scr02e
	scr031, scr032, scr033, scr034, scr035, scr036
	scr04a, scr04b1, scr04b2, scr04b3, scr04b4, scr04c, scr04d1, scr04d2, scr04e, scr04f1, scr04f2, scr04f3, scr04g, scr04h
	scr05
	scr06a, scr06b1, scr06b2, scr06c
	scr7
	scr08a1, scr08a2, scr08a3, scr08b
	scr09a11, scr09a12, scr09a21, scr09a22, scr09a31, scr09a32, scr09a41, scr09a42, scr09a51, scr09a52, scr09a53, scr09b1, scr09b2, scr09b3, scr09b4, scr09b5, scr09b6, scr09c1, scr09c2, scr09d01, scr09d02, scr09d03, scr09d04, scr09d05, scr09d06, scr09d07, scr09d08, scr09d09, scr09d10, scr09d11, scr09d12, scr09d13, scr09e1, scr09e2, scr09f1, scr09f2, scr09g1, scr09g2, scr09h1, scr09h2, scr09j, scr09k, scr09l
	scr100, scr10a, scr10b, scr10c, scr10d
	scr110, scr11a, scr11b
	scr120, scr12a, scr12b, scr12c, scr12d
	scr130, scr13a, scr13b
	scr14a1, scr14a2, scr14a3, scr14a4, scr14b1, scr14b2, scr14b3, scr14b4
	scr15
COMMUNE [156 obs.]	cmt00a, cmt00b, cmt00c
	cmt011, cmt012
	cmt021, cmt022
	cmt031, cmt032, cmt033, cmt034, cmt035
	cmt0400, cmt0401, cmt0402, cmt0403, cmt0404, cmt0405, cmt0406, cmt0407, cmt0408, cmt0409, cmt0410, cmt0411
	cmt05
	cmt061, cmt062, cmt063
	cmt071, cmt072, cmt073
	cmt081, cmt082, cmt083, cmt084
	cmt091, cmt092
SCHOOL [574 obs.]	cmt101,cmt102, cmt103, cmt104, cmt105, cmt106, cmt107, cmt108, cmt109, cmt110, cmt111, cmt112
CLINIC [155 obs.]	cmt21, cmt22, cmt23, cmt24, cmt25
PRICE [190 obs.]	cmt30, cmt31, cmt32, cmt33, cmt34

Household Data

Information collection for this part is based on the household questionnaire in which questions are grouped into 15 sections. Data on household make up 115 files arranged in order compatible with the structure of the questionnaire.

There were 2 types of observations in the data set

- Data on household members (section 1, 2, 3, 4, 5 and 15)
- Data on the entire household (section 0, 6, 7, 9, 10, 11, 12, 13 and 14)
- Data on one married woman age 15-49 selected randomly in each household (section 8)

Name of files in the data set

The names of household data files are generally in the form of **scrnnsspp**.

Of which:

1. Names of files begin with 3 characters **scr** (screen)
2. **nn** indicates section numbers, which can take values from 00 to 15 as in the questionnaire.
3. **ss** indicates the name of parts of sections if sections consist of 2 parts and above that include 1 or 2 characters (such as A, B, C, D or A1, A2,...); These two characters aren't used if sections have no sub-parts.
4. **pp** indicates the order of pages of parts (or sections) in the questionnaire, if the part (or the section) includes more than 1 screen that usually correspond to 1 page of the questionnaire then 1 or 2 characters are used; These two characters aren't used if the part (or the section) only has 1 screen.

For example:

- Section 0 - SURVEY INFORMATION consists of 4 parts A, B, C, D and each part placed at 1 pages in the questionnaire (as corresponding screens) so names of 4 files are **scr00a**, **scr00b**, **scr00c**, **scr00d** respectively.
- Section 1, part A consists of 2 screens so names of these files are **scr01a1**, **scr01a2** respectively.
- Section 9, part A1 consists of 2 screens (from 1 pages in the questionnaire) so names of these files are **scr09a11**, **scr09a12** respectively.
- Section 9, part D consists of up to 13 screens so names of these files are **scr09d01**, **scr09d02**, **scr09d03**, ..., **scr09d10**, **scr09d11**, **scr09d12**, **scr09d13** respectively.

Name of variables in data files

1. There are 2 variables in every household data file: **cluster** and **househol**
 - Variable **cluster** indicates the code of the cluster (or village) where the household lives.
 - Variable **househol** indicates the code of the household.
2. In individual level files there is a variable named **idcode** which is the individual identification code.
3. Variable **vlsscode** in the file named **scr01d** indicates the member code of households interviewed in Vietnam Living Standards Survey 1992-93.
4. The name of variables corresponding to questions in the questionnaire is in the form of **snnppqyyx**, where:
 - Name of variable begin with character **s**
 - **nn** indicates section numbers, which can take values from 00 to 15.
 - **pp** indicates the name of parts of sections, that can take values as a, b, c,..., h, i, j, k, l, or a1, a2, a3,...; b1, b2, b3, etc.
 - **qyy** is the question number in each part, each section of the questionnaire (where **q** is fixed and **yy** is the question number).
 - **x** is one or two letters or numbers to clarify the question, for example: **y** for question related to year, **m** for question related to month or minutes, **d** for question related to day (or distance), **h** related to hours, **c** for code, **q** is for quantity, **u** for question related to unit, **t** for question related to number of times, **a** for question related to age (or amount of money), **cc** for question related to crop codes, **fc** for question related to food codes, **f**, **s**, **t** indicate first, second, third (for questions allowing more than one answer)

In addition, some other names are used such as **line** for crop number, **ic** for item codes, **en** for enterprise number, etc.

For example:

- Section 1 A, question 5 related to year would be **s1aq05y**
- Section 4 B, question 22 related to lunch bonuses would be **s4bq22b**
- Section 12C, question **s12cln** refers to the line number next to question 1, while **s12cq01** refers to the code of the type of durable good seen in the table to the left.

Data on Community, School, Commune Health Center and Prices

The names of files are generally in the form of **cmtnnpp**.

Of which:

1. Name of files begin with 3 characters **cmt**
2. **nn** indicates section numbers, which can take values from 00 to 09.
3. **pp** indicates the order of entry screens, which correspond to 1 part (or 1 section), 1 page of the commune questionnaire or group of questions in 1 part (or 1 section).

For example:

- Section 0 - SURVEY INFORMATION consists of 3 parts a, b, c corresponding to 3 files: **cmt00a**, **cmt00b**, **cmt00c**.
- Section 1 - Basic Physical and demographic characteristics of the commune/ward and selected villages/blocks consists of 2 files: **cmt011** that include 3 questions: question 1, 2 and 3; **cmt012** that include question 4 to 8.

In general, names of variables in commune data files are the same as names of variables in household data files. However, some special variables are used such as: **lsp** for types of product in file **cmt031**; **idc** for factory code in file **cmt033**; **ccod** for village code in file **cmt061** and **cmt062**, etc. In addition, some special variables aren't set up according to the above principle such as: **ccc** for health equipment code in file **cmt23**, but **ccc** in file **cmt24** is medicine code.

Appendix C. List of Related Documents

The following documents can be obtained from the World Bank, Living Standards Measurement Study, Development Research Group (DECRG). They can be downloaded from the LSMS web site:

<http://www.worldbank.org/lsms/lsmshome.html>

Users may also request copies from:

Living Standards Measurement Study
World Bank
1818 H Street, NW
MSN MC3-306
Washington, DC 20433

A. Questionnaires

Household Questionnaire (English & Vietnamese versions), 158 pages including all facing pages with codes.

Community Questionnaire (English & Vietnamese versions), 33 pages

School Questionnaire (English & Vietnamese versions), 4 pages

Commune Health Center Questionnaire (English & Vietnamese versions), 3 pages

Price Questionnaire (English & Vietnamese versions), 5 pages

B. Training manuals

Interviewer Manual (English & Vietnamese), 105 pages

Anthropometry and price questionnaire manual (English & Vietnamese), 47 pages

Supervisor Manual (English & Vietnamese), 53 pages

Manual for Commune, School and Commune Health Center Questionnaires (English & Vietnamese), 47 pages

Supplementary instructions to survey teams (English & Vietnamese), 10 pages

C. Other

Vietnam Living Standards Survey: Recommended Sample Design, July 1997, by Chris Scott. (English only) 20 pages

Household Survey Data dictionary (with Vietnamese labels), 33 pages

Commune, School, Health Clinic, and Price Survey Data dictionary (with Vietnamese labels) 22 pages

Means and Standard Deviations of variables in the household survey, 53 pages

Spreadsheet used for calculating weights

Appendix D. Price Indices

In calculating items such as income, wages or expenditures in real terms there are two types of price indices.

- monthly price index over the period of the survey to adjust to January 1998 prices
- regional price deflators to adjust regional and urban/rural price differences to a national average

For each of these indices, we also have food and non-food price indices, and for monthly price indices a rice and non-rice food price index along with the overall index. All price indices (5 monthly and 3 regional) are included in the hhexp98n.dta files which are explained in Appendix E.

Monthly price indices

The month on month price changes are from the Vietnamese CPI which calculates a weighted basket of food and non-food items, as well as an overall CPI, and is available for the whole country over the entire period encompassing the first and second VLSS surveys. Data are from both urban and rural locations from all provinces in the country. The consumption basket weights used are from the 1995 MPHS survey.

The monthly price deflators are called mcpif, mcpir, mcpinrf, mcpinf and mcpi for food, rice, non-rice food, nonfood and overall prices respectively. The mcpi are set with January 1998 = 1, and each month of the survey, either in 1992 or 1998 is set in relation to this date. The ratio between January 1998 and January 1992 is 1.456 which is the factor one could use to multiply January 1992 figures to adjust them to January 1998 equivalent prices.

Note, for holiday food expenditures, as they occurred in January are deflated as follows: Holiday expenditures reported prior to Tet 1998 are expenditures for Tet in 1997. These expenditures are inflated to January 1998 prices using the GSO food price index.

Regional Price indices

For 1998, the GSO created regional price indices based on their normal CPI data, but disaggregated by region and urban/rural for purposes of the 1997-98 VLSS survey. They also include food, non-food and overall price indices. An explanation as written by the Price specialists of the GSO is translated and attached as an annex to this Appendix.

The regional deflators are called rcpi, rcpinf, rcpi (originally called inprice) for food, nonfood and overall prices respectively. The rcpi, rcpinf and rcpi are set with the national average being 1. Because of slight weighting differences between the CPI and the VLSS, the regional indices supplied by the GSO were normalized so the population weighted average for the nation equaled 1 which required only slight adjustments.

To adjust for regional price differences, just divide the nominal figure by the price index.

Table D1: Monthly Price deflators

Year	Month	Deflator				
		General	Non-food	Food	Rice	Non-rice Food
1997	December	98.40	99.22	97.94	98.53	97.65
1998	January	100.00	100.00	100.00	100.00	100.00
	February	102.20	100.87	103.00	103.75	102.63
	March	101.40	99.77	102.18	104.29	101.13
	April	103.00	100.10	104.63	114.56	99.71
	May	104.50	100.42	106.83	120.96	99.83
	June	104.50	100.45	106.72	117.81	101.23
	July	103.90	100.56	105.86	116.74	100.48
	August	105.10	101.06	107.35	121.76	100.21
	September	106.10	101.71	108.53	126.51	99.62
	October	106.40	101.87	108.85	125.34	100.69
	November	106.50	102.43	108.64	125.08	100.50
	December	107.40	102.67	110.05	125.61	102.34

Table D2: Regional Price deflators

	Overall		Non-food		Food	
	Rural	Urban	Rural	Urban	Rural	Urban
Northern Uplands	101.78	99.30	104.58	98.82	100.21	99.56
Red River Delta	91.62	105.60	92.19	101.71	91.30	107.78
North Central Coast	93.80	102.50	94.97	97.33	93.14	105.40
South Central Coast	97.42	105.30	97.77	95.72	97.22	110.67
Central Highlands	106.02	..	102.03	..	108.26	..
Southeast	96.54	113.42	96.87	105.21	96.36	118.02
Mekong Delta	102.69	101.33	105.02	105.30	101.39	99.11

Annex: Summary explanation for calculating regional consumer price indices compared to national price levels 1998

Trade and Prices Department
General Statistical Office

Regional price indices

1. Calculate the average price for each province, separately for urban and rural areas in 1998

The average 1998 price for each province is the simple average of the price for each item for the 12 months of the year.

$$\bar{P}_{Tj} = \frac{\sum_{t=1}^{12} P_{Tjt}}{12}$$

Where:

\bar{P}_{Tj} = average price in 1998 for item j

P_{Tjt} = price in month t for item j in the province

12 = 12 months in the year 1998

t = month

T = province

j = priced item

2. Calculate the regional and urban/rural average price for 1998

The simple average of the provincial prices as calculated above for each region and urban/rural location.

$$\bar{P}_{Vj} = \frac{\sum_{T=1}^{n_V} \bar{P}_{Tj}}{n_V}$$

Where:

\bar{P}_{Vj} = average price for region (separately for urban/rural) V for item j (v=1-14), (j=1,...,m)

\bar{P}_{Tj} = average price in 1998 for item j in province T (tỉnh in Vietnamese is province)

n_V = number of provinces in region V

m = total number of items for which average prices are calculated

V = region

3. Calculate general average price for the whole country for each item

$$\bar{P}_{Cj} = \frac{\sum_{t=1}^{14} \bar{P}_{Vj}}{V}$$

Where:

\bar{P}_{Cj} = average price in the whole country for item j

\bar{P}_{Vj} = average price in 1998 for item j in region V

V = number of regions used to calculate average prices for the whole country (vung is region in Vietnamese)

4. Calculate the price index for foodstuffs and non-food items and overall price index for urban and rural areas separately

Using the Laspeyres price index formula; In which, prices in the original period are the average prices of the whole country in 1998 (calculated in 3) , and fixed weights being the structure of consumption for the whole country in 1995 (results from the MPHS)to calculate for 7 regions and urban/rural areas.

$$Ip_v = \frac{\sum_{g=1}^G ip_{gv} \cdot w_{g95}}{\sum_{n=1}^N w_{g95}}$$

Where:

Ip_v = price index for each region

ip_{gv} =price index for each item representing a group (g) of prices (exp. items) in region v

w_{g95} =structure of expenditures (price times amount) for the whole country for ip_{gv} (general structure used to calculate all 3 indices (food, non-food, overall))

5. Calculate the overall price index (for combined urban and rural areas, for each of the three price indices).

Using weighted averages, for which the weight is the population structure between urban and rural areas in 1997 in order to combine the urban and rural indices above.

$$Ip_c = \frac{\sum_{ds=1}^2 Ip_v \cdot w_{ds97}}{\sum_{ds=1}^2 w_{ds97}}$$

Where:

Ip_c = general price index for each region (1-7)

ip_v = price index calculated in section 4

w_{ds97} = structure of population by urban/rural residence in 1997

ds=urban or rural indicator

$$\sum w_{ds97} = 100$$

Appendix E. Household Expenditure Calculations

In arriving at a measure of household consumption expenditures, the objective is to include, as accurately as possible, all expenditures that contributed to the enhancement of household welfare in the year preceding the interview. The VLSS contains expenditure information on various food and non-food items. The way in which this information is used to arrive at a measure of household expenditures in the 12 months preceding the interview is described in this Appendix.

Food expenditures

Household expenditures on food items are available in sections 11A and 11B. In Vietnam, the lunar new year (Tet) causes a considerable departure from normal spending patterns. Food expenditure information during Tet and other holidays (hereafter referred to as just Tet) is, therefore collected separately in section 11A. Non-holiday expenditure information is available in section 11B.

Tet expenditures cover about a two-week period. For 18 food items/categories, information is available on the quantity and value of market purchases. For these same foods, information is also obtained on the quantity consumed out of home production, and the value of this quantity if it had been purchased in the market. The data on value of market purchases and value of home production consumed by the household is to be included in household expenditures: the procedure is described later.

Value of annual market purchases

Section 11B contains non-holiday food expenditure information on 45 food items. Data is obtained on value of purchases made since the first interview, quantity and value of market purchases in the preceding 12 months, and quantity and value of foods consumed out of home production in 12 months. For obtaining annual food expenditures, the information on value of purchases made in the prior 2 weeks (since the first interview) is ignored.

Market purchase information is obtained as follows. The respondents are asked for the number of months during which market purchases were made. They are then asked for the number of purchases made during these months. To allow for differences in purchase frequency, both across food items and across respondents, respondents can provide purchase frequency information for one of 6 different recall periods separately for each item. Number of purchases can therefore be reported as number of purchases per day, week, month, quarter, half-year or year. For example, a respondent could report that during 6 months of the year bananas were purchased twice a week and that during the whole year, duck was purchased once per month. Respondents then provide the information on the average quantity and value of a typical purchase. It is this information – months of purchase, number of purchases, recall period and value of each purchase – that is used to obtain the value of non-holiday market food expenditures in the 12 months preceding the interview.

The procedure used to obtain annual non-holiday market expenditures consists of the following steps.

- Step 1 converts purchase frequency information – number of purchases and recall periods – to a single “purchases per month” variable for all foods and all households.
- Step 2, the variable for months of purchase is scaled down if a household has Tet expenditures for that food item. Assuming that the Tet and holiday period covers 2 weeks or half a month, months of purchase are multiplied by 11.5/12 if there are Tet expenditures for a food item. No adjustments are made if the recall time period exceeds a month, i.e. if purchase frequency is reported as number of times per quarter, half-year or year.

- In step 3, the “purchases per month” variable is multiplied by variables for number of months of purchase and the value of each purchase to obtain the value of annual non-holiday market purchases.

Value of home produce consumed during the year.

Information on home produce consumed by the household is obtained as follows. Respondents are asked for the number of months during which home produce is consumed, the total quantity consumed (during these months) and the value of such consumption if it was to be purchased in the market. Calculating annual expenditures on home production is relatively simple.

If a household does not consume anything from home production for a particular food item during Tet, the value of annual expenditures on home production is simply equal to the value provided by the respondent. On the other hand, if a household reports Tet home production expenditures for a food item, one needs to scale down the values reported in section 11B. The rule used to scale down section 11B expenditures is as follows.

- (i) If months of consumption, from home production, are greater than 6, expenditures are scaled down so that they cover 11 and a half months – i.e. value is multiplied by 11.5/12.
- (ii) If months of consumption are 6 or less, then no adjustment is made to expenditures.

Non-food expenditures

Non-food expenditures include expenditures on clothing, footwear, personal care, entertainment and recreation, transportation, housing supplies, housing, furniture, household appliances, other consumer durables, education and health. These are discussed in the order in which they are added to food expenditures, to obtain total household expenditures.

Daily expenses

Section 12A contains information on 15 frequently purchased non-food items (plus “other”). This section also includes non-purchase of these same items including home production or receipt of gifts. Expenditures items include areca and betel, children’s pocket money, cooking fuel, gasoline, matches or lighters, soap and detergent, toiletries, flowers, parking fees, lottery tickets and worship items purchased in the 4 weeks preceding the interview. To get annual expenditures, these expenditures over a 4 week reference period are multiplied by thirteen. The one item excluded from our calculation of household expenditure is lottery tickets (113). These expenses are excluded because they enhance well-being only when a household wins the lottery, and in that case the amount won is likely to be reflected in increased expenditures on other goods.

Differences with 1992-93

In 1992-93 home production/ non-purchase was not asked about for non-food items, but it was in 1997-98 for coal, wood, sawdust, chaff and flowers.
--

Annual expenditures

Less frequent non-food expenditure information is obtained in section 12B. Information is obtained on expenditure in the 2 weeks since the first interview, and both purchase and non-purchase receipt (home production or gift) of each item in the last 12 months. This section includes expenditures on fabric, ready-made clothing, blankets, footwear, tailoring services, pots and pans, household supplies, housing materials, motorcycle, bicycle and car repairs and spare parts, public transportation fees, books and stationery, recreation, taxes, etc. Since annual expenditure is available for each item, no calculation is needed for items in this section.

There are four types of expenditure items for which information was collected, but not utilized in creating expenditure variables. The first includes items that do not increase economic well-being and includes (item number in parentheses): taxes (155), security fund (156), labor contribution

(157), social funds (158), insurance (159), and gift transfer (162) which bring benefits to others, not to the household.

The second type of item is those for which expenditures are likely to be reflected in the current value of the relevant consumer durable. These include: other household items (132) which include a large number of high price items which are likely to be consumer durables, bike tubes and tires(136), bicycle spare parts (137), motorbike, car tires, tubes (138), motorcycle, car spare parts (139), Maintenance and repair of bicycle or motorbike or car (140).

The third type are those related to house building or repairs which are included in the use value of owned housing including: construction materials (133), expenses for building new house (134) and home repairs, painting, etc.(135).

The fourth type are those considered to be savings or investment rather than consumption including gold, silver, precious gems for saving (150) and contributions to informal credit arrangements, purchase of stocks or bonds (164).

Consumer durables

Consumer durables provide a flow of services over a period of time, and therefore it is misleading to include annual expenditures on them in a measure of household expenditure that is to serve as a measure of a household's standard of living. Excluding those expenditures is also inappropriate since they do contribute to an improvement in a household's economic well-being. The objective, therefore, is to obtain a measure of the value of services provided by consumer durables – their “use value” in the 12 months preceding the interview.

To obtain a measure of “use value” what is needed is the price at which the durable was purchased, the duration for which the household has been in possession of the item, and the durable's current value. Based on these three variables, along with a price index for deriving real values for purchase price and current value, one can construct a measure of the monetary value of services derived from consumer durables. These data are available from the household survey and are utilized in the following manner to obtain use values.

The first step consists of deriving depreciation rates for each type of consumer durable. These are derived as follows. The relationship between the real value of a good at the time it was bought (VB) and its value at the time of the interview (VT) is:

$$VT = VB(1 - d)^t$$

where t is the number of years since purchase, and d is the depreciation rate. Since the survey collects information on purchase price, resale value at the time of the interview, and year of purchase, it is possible to derive the implicit depreciation rate for each type of consumer durable for each household. Specifically, the depreciation rate can be derived from the formula above as equal to:

$$d = 1 - \left(\frac{VT}{VB} \right)^{1/t}$$

To minimize the influence of errors and biases in self-reported resale values and purchase prices, the median depreciation rate for each type of consumer durable is used.

These depreciation rates were calculated using only consumer durables acquired after 1986. This was done because reasonable price indices for earlier years, which are necessary for converting nominal purchase prices to real values, do not exist.

The second step consists of using the depreciation rate along with the real interest rate to obtain the annual use value of a durable good, in order to measure the opportunity cost to the owner of using the good for one year, instead of selling it at the beginning of the year and investing the money at the real interest rate. Therefore, use value is obtained as:

$$\text{Use value} = VI(1+d) \times (d+0.05)$$

where 0.05 is an estimate of the real interest rate and $VI(1+d)$ is the value of the consumer durable 12 months before the interview.

Utilities

Section 6 contains information on housing characteristics and housing expenses. Expenditure data on electricity, drinking water, laundry and bathing water are obtained with the possibility of variable reporting periods (similar to those for food expenditures). Similar information is also obtained for household payments for garbage disposal services. For all these expenditures, the reported expenditure information is scaled up (if reporting period is less than a year) to obtain annual expenditures for electricity, drinking water, laundry and bathing water, and garbage disposal.

Rent

For households who rent their dwelling, rental expenditures for the year are clearly their housing expenses. For households who live in dwellings they own, the true cost of living in the owned dwelling is not zero, but the opportunity cost of living in that dwelling. If there is a competitive rental market for dwellings, then the rental amount paid by households who rent their dwelling is likely to be an accurate measure of the opportunity cost of living in similar dwellings, and this information can be utilized to impute a rent for those who live in dwellings they own. The usual procedure, in such cases, is to run a regression of rental values on housing characteristics and then use the coefficients from such a regression to impute rental values for those who do not rent their dwellings.

Unfortunately, in the Vietnamese case housing rental markets are still severely underdeveloped. In 1997-98 only 278 households rent or borrow their dwellings, 23 of whom rent from the private sector and only 12 of whom have to pay rent. This leaves us with too little information to calculate rental values. An additional complication is that between 1992-93 and 1997-98, a Land Law was passed and put into effect which led to rapid increases in the households' appraisal of the value of their dwelling (and the land on which it rests.) Use of some fixed proportion of the value of the dwelling would lead to very different use values in the two years for the same dwelling, whether or not it had undergone any repairs or upgrading. A rather arbitrary procedure was utilized to provide some estimate of the housing value which could be comparable between the two years.

Using information from 1992-93, the ratio of 3% of the predicted house value in relation to total non-housing, non-food expenditures for the third quintile was calculated separately for urban and rural households using current prices. This same ratio was then applied to the 1998 current non-housing, non-food expenditures to give nominal housing use value in 1997-98. The underlying assumptions are that 3% is a good estimate of the use value in 1992-93, and that the ratio of housing use value to non-food expenditures has not changed between 1992-93 and 1997-98. Information on the value of housing in 1997-98 was not utilized in obtaining this estimate.

Education

Household expenditures on tuition, registration, uniforms, books, etc. for each household member are obtained in section 2D of the survey. The questionnaire also provides the totals for all such expenditures for each member. In addition to this, expenditures on foreign language and

computer courses, and any other courses such as driving, typing, accounting are available in section 2E for each household member.

Differences with 1992-93

In 1992-93, a question was asked about expenditures for foreign language training. In 1997-98, the question on expenditures for foreign language training included also computer training making the responses non-comparable. In addition, in 1997-98, an additional question was asked about any expenditures on other training courses which was not asked in 1992-93. Otherwise, it is assumed that any extra items asked in 1998 were included in "other" in 1992-93 and vice versa.

Health

Section 3 contains information on health expenditures during the past 4 weeks for various providers. Information is also available on hospital expenditures in the past 12 months, on total expenditures in the past 12 months and on insurance expenditures in the past 12 months. This makes it possible to calculate total health expenditures in 2 ways. First, one could take the estimates of expenditures in the past 4 weeks and multiply by thirteen, add in hospital and insurance expenditures in the past 12 months. Or alternatively, take total health expenditures in the past 12 months and add to hospital and insurance expenditures in the past 12 months.

Differences with 1992-93

In 1992-93 one question asked about total health expenditures in the past 12 months which was utilized in calculating total expenditures. Questions related to health seeking in the past 4 weeks were not comprehensive and could not be used to estimate total health expenditures. No question asked about health insurance costs.

In 1997-98, a complete set of questions referring to health expenditures in the past 4 weeks plus information on hospitalization and insurance costs allow one to calculate an estimate of total health expenditures in the past year which can be utilized in aggregate analysis. This assumes that the health expenditures of the individuals sick in the past 4 weeks are on average comparable to the expenditures of people sick in the other 11 months of the year. When added up and divided by the total number of people, one gets a good estimate of average health expenditures for different groups of people. However, for individual level analysis such as in regressions, one must still utilize information from the question on total health expenditures in the past 12 weeks. Estimates using information from the past 4 weeks tend to be higher than for those in the past 12 weeks as people recall more in a shorter period of time.
--

Tobacco expenditures

The last part of section 3 in the 1997-98 VLSS asks each individual about his/her tobacco spending habits over a 12 month reference period and including consumption of home production. A tobacco expenditure variable is calculated by adding up spending on all forms of tobacco in the past 12 months, plus the value of home produced tobacco consumed for all individuals in the household.

Differences with 1992-93

In 1992-93 a household level question was asked about tobacco expenditures. As a representative of the household responding about non-food consumption could not know the actual consumption as accurately as each individual it is expected that in 1992-93 the VLSS undercounted tobacco expenditures and the two ways of measuring tobacco expenditures are not comparable.
--

Data problems and Corrections

In theory, a variable-recall procedure, such as the one used in the VLSS, should provide more accurate expenditure information than a constant-recall procedure such as asking for expenditure

information for the month preceding the date of the interview. But a variable-recall procedure is also more demanding on the part of the interviewer and the respondent. And from an analytical perspective too, it requires checking the data to see if the various parts are consistent with each other. In constructing the expenditure variable the following checks were performed for outliers:

- (1) Per capita consumption of various food items
- (2) The distributions of all component variables for food expenditure data – months of purchase, number of purchases, recall period time units, quantity of each purchase, quantity units and value of each purchase – were examined to identify incorrect values and outlying cases.
- (3) Unit prices for food items were calculated by dividing expenditure values for food items by quantities, and their distributions were checked for outlying values. The advantage of checking distributions of unit prices is that both tails of the distribution can be checked for outlying values; in the case of quantity and value distributions one can only check the upper tail of the distribution because any small value is reasonable. In checking these distributions, comparisons were made with data from the price questionnaire.
- (4) Browsing through the data to explore distributions and patterns beyond the information provided by descriptive statistics such as means and order statistics.

Variables in the household expenditure file made from the Vietnam Living Standards Survey 1997-98

Household expenditure variables are calculated for the various expenditure items described above for the Vietnam Living Standards Survey 1997-98.

Because of differences for some items between 1997-98 and 1992-93 surveys including: expenditure on cigarette and tobacco; health insurance; foreign language, computer skills, vocational training; and home production of coal, wood, sawdust, chaff, two variables for expenditures are created, one for comparison with 1992-93 and one which would be considered a total expenditure variable for the 1997-98 cross-section data.

In the household expenditure file for year 1997-98 (**hhexp98n.dta**), if the name of a variable ends with “1” then this variable can be compared to VLSS92-93; if the name of a variable ends with “2” then this variable is the best variable for VLSS97-98. Variables without “1” or “2” at the end are variables without any difference in the list of expenditure items between the two surveys. Some variables calculated from new or different expenditure items in 1997-98 may not exist in the 1992-93 expenditure files such as: **langcomp**; **otheredu**; **insrx12m**. Variable **quint98** can be compared to variable **quint92**; but variable **quint98b** is the best variable for VLSS97-98.

Names and meaning of variables in the household expenditure file of the Vietnam Living Standards Survey 1997-98 are as follows:

File name: **hhexp98n.dta**

Total of observation: 5,999 households

Total of variable: 67 variables

Size of file: 1,553,741 bytes

Order and name of variable	Meaning
1. household	Household code
2. year	Year of interview
3. month	Month of interview
4. vlssmphs	Code 1: VLSS92-93 cluster, Code 2: MPHS cluster
5. sex	Sex of head of household (1: Male; 2: Female).

6. age	Age of head of household
7. agegroup	Age group of head of household (under 19; from 20 to 29; from 30 to 39; from 40 to 49; from 50 to 59; from 60 to 69; 70 and above.
8. comped98	Highest diploma obtained by head of household
9. educyr98	Number of schooling years of head of household
10. farm	Classify households into agricultural and non-agricultural household (1: agricultural households; 0: non-agricultural households)
11. urban98	Code of urban/rural for year 1998 (1: urban; 0: rural) (some changes since 1992-93)
12. urban92	Code of urban/rural for year 1992 (1: urban; 0: rural)
13. province	Code of Province/city (according to current government classification of administrative units, not the codes in the questionnaire)
14. reg7	Regional code for 7 regions (classification in 1992-93)
15. reg8	Regional code for 8 regions (current classification)
16. reg10	Regional code for 10 regions (current classification)
17. hhsiz	Size of household
18. hhcat	Classify households by household size (values: 1 to 10 corresponding to the household with 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 persons and above).
19. wt	Household weight (number of households represented by 1 surveyed household)
20. hhsizewt	Household size weight (number of persons represented by 1 surveyed household)
21. vill	Code of village
22. reg10o	Regional code for 10 regions when sample selection was done. It is used for the strata when we run SVY commands. It is different from the variable reg10 which contains the updated and current classification of clusters by their urban/rural and regional status.
23. commune	Code of commune/ward. It is used as the PSU when we run SVY commands
24. ricexpd	Expenditure on rice of surveyed households at current price
25. nonrice	Expenditure on food and foodstuff excluding rice of surveyed households at current price
26. totnfdx1	Expenditure on non-food purchased or bartered for by surveyed households at current price (comparable to 1993 figures)
27. totnfdhp	Value of home production of all non-food items consumed by surveyed households at current price (not comparable to 1993 figures)
28. totnfdh1	Value of home production of all non-food items excluding coal, wood, sawdust, chaff consumed by surveyed households at current price (comparable to 1993)
29. langcomp	Expenditure on foreign language and computer skills training (not comparable to 1993)
30. otheredu	Expenditure on other training courses such as: typing, accounting, car driving, etc. (not comparable to 1993)
31. educnexp	Expenditure on education (comparable to 1993)
32. insrx12m	Expenditure on health insurance (not comparable with 1993)
33. hhexp12m	Expenditure on health (in the past 4 weeks and 12 months) (not comparable with 1993)
34. hhexp12i	Expenditure on health (only in the past 12 months) excluding expenditure on health insurance (comparable with 1993)
35. hhexp12i	Expenditure on health (only in the past 12 months) including expenditure on health insurance (not comparable with 1993)

36. durbuser	Utilization value of durable goods of surveyed households (comparable with 1993)
37. garbexp	Expenditure on garbage disposal (comparable with 1993)
38. waterexp	Expenditure on water (comparable with 1993)
39. elecexp	Expenditure on electricity (comparable with 1993)
40. tobacco	Expenditure on cigarette and tobacco (not comparable with 1993)
41. rentexp3	Estimated value of house use (comparable with 1993)
42. mcpif	Food and foodstuff price index (January, 1998=1)
43. mcpinf	Non-food and non-foodstuff price index (January, 1998=1)
44. mcpi	Goods price index (January, 1998=1)
45. rcpi	Food and foodstuff price index in 1998 by region
46. rcpinf	Non-food and non-foodstuff price index in 1998 by region
47. rcpi	Goods price index in 1992 by region
48. mcpir	Rice price index (January, 1993=1)
49. mcpinrf	Food and foodstuff price index excluding rice (January, 1993=1)
50. food	Expenditure on food and foodstuff at current price (comparable with 1993)
51. nonfood1	Expenditure on non-food and non-foodstuff at current price (including expenditure on health; education; value of house use; electricity, water; garbage disposal and others <i>but excluding expenditure on foreign language, computer skills training</i> (langcomp); <i>other training courses</i> (otheredu); <i>tobacco and cigarette</i> (tobaco); <i>health insurance</i> (insrx12m); <i>home production of coal, wood, sawdust, chaff</i>) (comparable to 1993)
52. hhexp1	Total expenditure at current price (food+nonfood1) (comparable to 1993)
53. pcexp1	Expenditure per capita of surveyed household at current price (hhexp1/hhsize) (comparable to 1993)
54. pcfdex1	Expenditure on food and foodstuff per capita of surveyed household at current price (food/hhsize) (comparable with 1993)
55. rlfood	Expenditure on food and foodstuff of surveyed household readjusted by price indexes of regions and months (January, 1998=1) (comparable with 1993)
56. rlnfd1	Expenditure on non-food and non-foodstuff of surveyed household readjusted by price indexes of regions and months (January, 1998=1) (comparable to 1993)
57. rlhhex1	Total expenditure readjusted by price indexes of regions and months (January, 1998=1) (comparable to 1993)
58. rlpce1	Expenditure per capita of surveyed household readjusted by price indexes of regions and months (January, 1998=1) (comparable to 1993)
59. rlpcedex	Expenditure on food and foodstuff per capita of surveyed household readjusted by price indexes of regions and months (January, 1998=1) (comparable with 1993)
60. quint98	Indicates which expenditure quintile group a household belongs to: quint98=1: poorest; quint98=2: poor-mid; quint98=3: middle; quint98=4: mid-upper; quint98=5: richest (comparable to 1993)
61. nonfood2	Expenditure on non-food and non-foodstuff at current price (including expenditure on health; education; value of house use; electricity, water; garbage disposal and others <i>but excluding expenditure on foreign language, computer skills training</i> (langcomp); <i>other training courses</i> (otheredu); <i>tobacco and cigarette</i> (tobaco); <i>health insurance</i> (insrx12m); <i>home production of coal, wood, sawdust, chaff</i>) (not comparable with 1993)
62. hhexp2	Total expenditure at current price (food+nonfood2). (not comparable with 1993)

63. pcexp2	Expenditure per capita of surveyed household at current price (hhexp2/hhsize). (not comparable with 1993)
64. rlnfd2	Expenditure on non-food and non-foodstuff (including all expenditure items collected in 1998) per capita of surveyed household readjusted by price indexes of regions and months (January, 1998=1) (not comparable with 1993)
65. rlhex2	Total household expenditure readjusted by price indexes of regions and months (January, 1998=1) (not comparable with 1993)
66. rlpex2	Expenditure per capita of surveyed household readjusted by price indexes of regions and months (January, 1998=1) (not comparable with 1993)
67. quint98b	Indicates which expenditure quintile group a household belongs to: quint98b=1: poorest; quint98b=2: poor-mid; quint98b=3: middle; quint98b=4: mid-upper; quint98b=5: richest (not comparable with 1993)

Appendix F. Region Codes

Correct region codes and province codes according to the questionnaire are found in the file longlat.dta and in longlatfin.xls.

10 Regions (Strata) of the Survey	
Code	Category
1	Major Urban Areas (Hanoi and Ho Chi Minh City)
2	Middle-Sized Urban Areas (Minor Cities)
3	Small Urban Areas (Provincial Towns)
4	Rural Northern Mountains and Midlands
5	Rural Red River Delta
6	Rural North Central Coast
7	Rural South Central Coast
8	Rural Central Highlands
9	Rural Southeast
10	Rural Mekong River Delta

Geophysical Regions (Commune Questionnaire)	
Code	Category
1	Coastal
2	Delta
3	Midlands
4	Low Mountains
5	High Mountains

Seven Administrative Regions (Used as basis for sampling and in 1992-93 survey)		
Code	Category	
1	Northern Mountains and Midlands	
	Bac Giang	Lao Cai
	Bac Kan	Phu Tho
	Bac Ninh	Quang Ninh
	Cao Bang	Son La
	Ha Giang	Thai Nguyen
	Hoa Binh	Tuyen Quang
	Lai Chau	Vinh Phuc
	Lang Son	Yen Bai
2	Red River Delta	
	Hai Duong	Hung Yen
	Hai Phong	Nam Dinh
	Ha Nam	Ninh Binh
	Hanoi	Thai Binh
	Ha Tay	

3	North Central Coast	
	Ha Tinh	Quang Tri
	Nghe An	Thanh Hoa
	Quang Binh	Thua Thien Hue
4	South Central Coast	
	Binh Dinh	Ninh Thuan
	Binh Thuan	Phu Yen
	Da Nang	Quang Nam
	Khanh Hoa	Quang Ngai
5	Central Highlands	
	Dak Lak	Kon Tum
	Gia Lai	Lam Dong
6	Southeast	
	Ba Ria Vung Tau	Dong Nai
	Binh Duong	Ho Chi Minh City
	Binh Phuoc	Tay Ninh
7	Mekong River Delta	
	An Giang	Kien Giang
	Bac Lieu	Long An
	Ben Tre	Soc Trang
	Ca Mau	Tien Giang
	Can Tho	Tra Vinh
	Dong Thap	Vinh Long

Eight Administrative Regions (Most recent classification)		
Cod e	Category	
1	Red River Delta	
	Hai Duong	Hung Yen
	Hai Phong	Nam Dinh
	Ha Nam	Ninh Binh
	Hanoi	Thai Binh
	Ha Tay	
2	Northeast	
	Bac Giang	Phu Tho
	Bac Kan	Quang Ninh
	Bac Ninh	Thai Nguyen
	Cao Bang	Tuyen Quang
	Ha Giang	Vinh Phuc
	Lang Son	Yen Bai
	Lao Cai	
3	Northwest	
	Hoa Binh	Son La
	Lai Chau	
4	North	

	Central Coast	
	Ha Tinh	Quang Tri
	Nghe An	Thanh Hoa
	Quang Binh	Thua Thien Hue
5	South Central Coast	
	Binh Dinh	Phu Yen
	Da Nang	Quang Nam
	Khanh Hoa	Quang Ngai
6	Central Highlands	
	Dak Lak (Dac Lac)	Gia Lai
	Kon Tum	
7	Southeast	
	Ba Ria Vung Tau	Ho Chi Minh City
	Binh Duong	Lam Dong
	Binh Thuan	Ninh Thuan
	Binh Phuoc	Tay Ninh
	Dong Nai	
8	Mekong River Delta	
	An Giang	Kien Giang
	Bac Lieu	Long An
	Ben Tre	Soc Trang
	Ca Mau	Tien Giang
	Can Tho	Tra Vinh
	Dong Thap	Vinh Long

Note: This is the most recent classification of provinces into regions by the GSO However, as the sampling was based on the 7 regions and 3 classifications of urban areas there may be insufficient observations in the Northwest and Central Highlands Regions for significant statistical results if the 8 regions are utilized in analysis

Appendix G. List of Communes, Districts and Provinces Sampled

File longlatfin.xls has the correct codes for communes, regions and provinces according to the VLSS questionnaire. The province code in Hhexp98n.dta corresponds to current administrative codes, but not to those listed in the questionnaire.

Names of clusters in the VLSS 1997-98			Commune
Province	District	Ward	Codes
Major Urban Areas			
Hanoi	Hoan Kiem	Hang Ma [®]	1
Hanoi	Dong Da	Trung Liet [®]	2
Hanoi	Dong Da	Thinh Quang	3
Hanoi	Tay Ho	Thuy Khue	4
Hanoi	Hai Ba Trung	Bach Mai	5
Hanoi	Cau Giay	Dich Vong+	6
Ho Chi Minh City	District 1	Pham Ngu Lao [®]	7
Ho Chi Minh City	District 1	Da Cao	8
Ho Chi Minh City	District 4	Ward 10 [®]	9
Ho Chi Minh City	District 5	Ward 13	10
Ho Chi Minh City	District 6	Ward 6 [®]	11
Ho Chi Minh City	District 6	Ward 1	12
Ho Chi Minh City	District 10	Ward 4 [®]	13
Ho Chi Minh City	District 10	Ward 5	14
Ho Chi Minh City	Tan Binh	Ward 20 [®]	15
Ho Chi Minh City	Tan Binh	Ward 18	16
Ho Chi Minh City	Go Vap	Ward 1 [®]	17
Ho Chi Minh City	Phu Nhuan	Ward 4 [®]	18
Ho Chi Minh City	Binh Thanh	Ward 5	19
Ho Chi Minh City	District 7 (Nha Be)	Binh Thuan+ [®]	20
Medium Urban Areas			
Thai Nguyen	Thai Nguyen City	Dong Quang [®]	21
Thai Nguyen	Thai Nguyen City	Tan Long	22
Hai Phong	Le Chan	Trai Cau [®]	23
Hai Phong	Hong Bang	Thuong Ly	24
Hai Phong	Do Son Town	Ngoc Hai	25
Nam Dinh	Nam Dinh City	Nang Tinh	26
Thanh Hoa	Thanh Hoa City	Ba Dinh [®]	27
Thanh Hoa	Thanh Hoa City	Dong Tho	28
Thua Thien Hue	Hue City	Phu Hiep [®]	29
Da Nang	Da Nang City	Hai Chau 2	30
Khanh Hoa	Nha Trang City	Vinh Truong [®]	31
Khanh Hoa	Nha Trang City	Phuong Sai	32
Dong Nai	Bien Hoa City	An Binh [®]	33
Dong Nai	Bien Hoa City	Tan Hoa	34
Tien Giang	My Tho City	Ward 1 [®]	35
Tien Giang	My Tho City	Ward 5	36
Minor Urban Areas			
Yen Bai	Yen Bai Town	Minh Tan [®]	37
Lao Cai	Cam Duong Town	Pom Han	38
Son La	Mai Son	To Hieu State Farm	39

Names of clusters in the VLSS 1997-98			Commune
Province	District	Ward	Codes
Bac Ninh	Bac Ninh Town	Ve An	40
Hung Yen	Tien Lu	Tien Lu Town*+	41
Ninh Binh	Ninh Binh Town	Van Giang®	42
Ninh Binh	Ninh Binh Town	Nam Thanh*+	43
Ninh Binh	Ninh Binh Town	Dong Thanh*+	44
Nghe An	Vinh City	Dong Vinh*+	45
Quang Nam	Dien Ban	Vinh Dien Town	46
Quang Ngai	Quang Ngai Town	Nghia Lo+	47
Binh Dinh	Quy Nhon City	Le Hong Phong	48
Khanh Hoa	Dien Khanh	Dien Khanh Town®	49
Binh Thuan	Ham Tan	La Gi Town	50
Binh Phuoc	Loc Ninh	Loc Ninh Town*+	51
Long An	Duc Hoa	Hiep Hoa Town	52
Dong Thap	Sa Dec Town	Ward 2®	53
Dong Thap	Sa Dec Town	Ward 3	54
Vinh Long	Tra On	Tra On Town®	55
Soc Trang	Soc Trang Town	Ward 4	56
Kien Giang	Rach Gia Town	Vinh Lac	57
Ca Mau	Thoi Binh	Thoi Binh Town	58
Rural Northern Mountains			
Ha Giang	Quan Ba	Quan Ba	59
Tuyen Quang	Yen Son	Tan Tien	60
Cao Bang	Hoa An	Dan Chu	61
Lang Son	Van Quan	Trang Phai	62
Yen Bai	Yen Binh	Bach Ha	63
Lao Cai	Muong Khuong	Cao Son	64
Thai Nguyen	Dong Hy	Tan Long	65
Son La	Mai Son	Chieng Mung	66
Phu Tho	Song Thao	Tien Luong	67
Phu Tho	Tam Thanh	Thuong Nong	68
Phu Tho	Thanh Son	Cu Dong	69
Vinh Phuc	Tam Dao	Ba Hien	70
Bac Giang	Luc Ngan	Kien Thanh	71
Bac Giang	Tan Yen	Lien Trung	72
Bac Giang	Luc Nam	Tam Di	73
Bac Ninh	Bac Ninh Town	Kinh Bac	74
Bac Ninh	Tien Son	Huong Mac	75
Bac Ninh	Thuan Thanh	Dinh To	76
Quang Ninh	Tien Yen	Hai Lang	77
Hoa Binh	Da Bac	Hao Ly	78
Hoa Binh	Kim Boi	Hop Thanh	79
Rural Red River Delta			
Hanoi	Dong Anh	Nam Hong	80
Hai Phong	Thuy Nguyen	Lien Khe	81
Hai Phong	Vinh Bao	Vinh An	82
Hung Yen	My Van	Nghia Tru	83
Hung Yen	Chau Giang	Van Phuc	84

Names of clusters in the VLSS 1997-98			Commune
Province	District	Ward	Codes
Hung Yen	Tien Lu	Ngo Quyen*	85
Hai Duong	Nam Sach	Ai Quoc	86
Hai Duong	Kinh Mon	Hiep Hoa Town	87
Hai Duong	Cam Giang	Tan Truong	88
Hai Duong	Thanh Mien	Doan Ket	89
Ha Tay	Chuong My	Van Vo	90
Ha Tay	Phu Xuyen	Phuong Duc	91
Ha Tay	Son Tay Town	Son Dong	92
Ha Tay	Ba Vi	Minh Quang	93
Ha Tay	Dan Phuong	Tan Hoi	94
Thai Binh	Dong Hung	Dong Vinh	95
Thai Binh	Vu Thu	Vu Doai	96
Thai Binh	Tien Hai	Dong Quy	97
Ha Nam	Duy Tien	Tien Hiep	98
Ha Nam	Binh Luc	An Lao	99
Ha Nam	Thanh Liem	Thanh Hai	100
Nam Dinh	Xuan Truong	Xuan Thuong	101
Nam Dinh	Nghia Hung	Nghia Chau	102
Nam Dinh	Hai Hau	Hai Ninh	103
Ninh Binh	Hoa Lu	Ninh Son	104
Rural North Central Coast			
Thanh Hoa	Trieu Son	Tho Phu	105
Thanh Hoa	Dong Son	Dong Van	106
Thanh Hoa	Ha Trung	Ha Giang	107
Thanh Hoa	Hoang Hoa	Hoang Son	108
Thanh Hoa	Quang Xuong	Quang Giao	109
Nghe An	Nghia Dan	Tay Hieu+	110
Nghe An	Vinh City	Hung Dong*	111
Nghe An	Quynh Luu	Quynh Yen	112
Nghe An	Nghi Loc	Nghi Dien	113
Nghe An	Do Luong	Nam Son	114
Nghe An	Anh Son	Lang Son	115
Nghe An	Tan Ky	Tan Ngai	116
Nghe An	Con Cuong	Bong Khe	117
Ha Tinh	Nghi Xuan	Xuan Hoi	118
Ha Tinh	Thach Ha	Thach Ban	119
Quang Binh	Bo Trach	Hung Trach	120
Quang Tri	Gio Linh	Gio Thanh	121
Thua Thien Hue	Huong Tra	Huong Toan	122
Thua Thien Hue	Phu Loc	Loc Binh	123
Rural South Central Coast			
Quang Nam	Nui Thanh	Tam My®	124
Quang Nam	Duy Xuyen	Duy An	125
Quang Nam	Tra My	Tra Dong	126
Da Nang	Hoa Vang	Hoa Qui	127
Quang Ngai	Son Tinh	Tinh Long®	128
Quang Ngai	Son Ha	Son Thuong	129

Names of clusters in the VLSS 1997-98			Commune
Province	District	Ward	Codes
Quang Ngai	Duc Pho	Pho Van	130
Binh Dinh	Phu My	My Hoa®	131
Binh Dinh	Hoai Nhon	Hoai Xuan	132
Binh Dinh	Tay Son	Binh Nghi	133
Phu Yen	Tuy Hoa Town	Hoa Tri	134
Khanh Hoa	Khanh Vinh	Cau Ba®	135
Khanh Hoa	Dien Khanh	Dien Phu	136
Ninh Thuan	Ninh Son	Lam Son	137
Binh Thuan	Ham Thuan Bac	Ham Thang®	138
Binh Thuan	Ham Thuan Bac	Ham Cuong	139
Rural Central Highlands			
Gia Lai	Kbang	Kroong®	140
Gia Lai	A Yun Pa	Chu Rang®	141
Gia Lai	Mang Giang	Kdang	142
Kon Tum	Dak Glei	Dak Choong®	143
Dak Lak	Krong Buk	Pong Drang®	144
Dak Lak	Krong Pak	Ea Kuang®	145
Dak Lak	Krong Bong	Hoa Le®	146
Dak Lak	Cu Zut	Hoa Phu	147
Dak Lak	Cu Mgar	Quang Phu	148
Lam Dong	Lam Ha	Phi To®	149
Lam Dong	Bao Lam	Loc Ngai®	150
Lam Dong	Di Linh	Tan Chau	151
Rural Southeast			
Ho Chi Minh City	Cu Chi	Trung Lap Thuong®	152
Ho Chi Minh City	Can Gio	Binh Khanh	153
Ho Chi Minh City	Hoc Mon	Tan Hiep	154
Binh Duong	Tan Uyen	Tan Binh®	155
Binh Duong	Ben Cat	An Tay	156
Binh Phuoc	Loc Ninh	Loc Thai*	157
Tay Ninh	Tan Bien	Tan Binh	158
Tay Ninh	Hoa Thanh	Truong Dong®	159
Tay Ninh	Duong Minh Chau	Truong Mit	160
Tay Ninh	Trang Bang	An Hoa	161
Dong Nai	Tan Phu	Phu Binh®	162
Dong Nai	Long Khanh	Xuan Duong®	163
Dong Nai	Vinh Cuu	Tan An	164
Dong Nai	Thong Nhat	Quang Trung	165
Dong Nai	Nhon Trach	Vinh Thanh	166
Baria Vung Tau	Chau Duc	Xuan Son®	167
Baria Vung Tau	Xuyen Moc	Binh Chau	168
Rural Mekong Delta			
Long An	Duc Hue	Binh Hoa Nam	169
Long An	Tan Tru	Binh Tinh	170
Dong Thap	Cao Lanh	My Xuong	171
Dong Thap	Chau Thanh	Phu Huu	172
An Giang	Phu Tan	Phu Hiep	173

Names of clusters in the VLSS 1997-98			Commune
Province	District	Ward	Codes
An Giang	Chau Phu	Thanh My Tay	174
An Giang	Thoai Son	Vong The	175
Tien Giang	Cai Lay	Thanh Loc	176
Tien Giang	Chau Thanh	Phu My	177
Tien Giang	Go Cong Dong	Tan Tay	178
Ben Tre	Mo Cay	Huong My	179
Ben Tre	Binh Dai	Tam Hiep	180
Vinh Long	Vinh Long Town	Tan Ngai	181
Vinh Long	Tam Binh	Binh Ninh	182
Tra Vinh	Cang Long	Nhi Long	183
Tra Vinh	Tra Cu	Ham Giang	184
Soc Trang	Thanh Tri	Chau Hung [*]	185
Soc Trang	Ke Sach	Thoi An Hoi	186
Soc Trang	Thanh Tri	Lam Kiet	187
Can Tho	O Mon	Truong Lac	188
Can Tho	Phung Hiep	Hiep Hung	189
Kien Giang	Giong Rieng	Thanh Hoa	190
Kien Giang	Vinh Thuan	Vinh Thuan	191
Ca Mau	Ca Mau Town	Hoa Thanh	192
Ca Mau	Dam Doi	Ta An Khuong	193
Bac Lieu	Gia Rai	Dinh Thanh	194
The following three clusters were included in 1992-93 but not in 1997-98			
Red River Delta			
Hai Phong	An Lao	My Duc	
Thai Binh	Quynh Phu	An Le	
Nam Dinh	Y Yen	Yen Phuc	

- Clusters marked with an asterix are those that have undergone some administrative changes such as villages splitting into two different communes and thus given separate cluster codes.
- Clusters marked with a + are those that have changed from urban to rural or rural to urban between 1992-93 and 1997-98 according to official government classifications.
- Clusters marked with a @ are those from the MPHS sample. Note that some of these clusters have only one village rather than two.

Appendix H. Occupation and Industrial Classifications Used in the VLSS 1997-98

OCCUPATIONAL CLASSIFICATION	
code	VLSS97-98
1	Leader in all fields and levels
2	Science and technology professionals
3	Health professionals
4	Education professionals
5	Other professionals
6	Personal services, protection and sales
7	Agriculture, Forestry and Fisheries
8	Skilled Manual Labor
9	Assemblers and Machine Operators
10	Unskilled workers

INDUSTRIAL CLASSIFICATION	
code	VLSS97-98
1	Agriculture, Forestry, Fisheries
2	Mining and Extracting
3	Manufacturing
4	Electricity and Water
5	Commerce
6	Transportat and Communication
7	Finance
8	Other Services Not listed below
9	Government Administration and National Defense
10	Education and Training
11	Health and Social Services
12	Culture and Sports
13	Communist Party and Mass organizations
14	Public Sanitation
15	Personal services
16	International organizations

Appendix I. List of Reports/papers Using the VLSS 1997-98

The following is a list of papers, published and unpublished, and research in progress that make use of the VLSS data. It is provided to guide researchers who are seeking to build on, but not reproduce work that has already been done.

Copies of all papers and publications resulting from the analysis of LSMS should be sent to the LSMS at:

Living Standards Measurement Study
Development Research Group
The World Bank
1818 H Street, NW
MSN MC3-306
Washington, D.C. 20433
USA

Joint Report of the Government of Vietnam, Donor, NGO Poverty Working Group. *Vietnam – Attacking Poverty*, Hanoi: World Bank Resident Mission. December 1999.

Bales, Sarah. *Vietnam's Labor Situation and Trends: Analysis based on the 1992-93 and 1997-98 Vietnam Living Standards Survey*. Mimeo. January, 2000.

General Statistical Office. *Statistical Abstract of the Vietnam Living Standards Survey 1997-98*. Hanoi: Statistical Publishing House. March, 2000.

Work in progress:

Nguyen Van Dong, Le Quy Dang, Ngo Doan Gac, Nguyen Ngoc Khoa, Hoang Van Kinh and Bob Baulch. *Determinants of Household Earned Income*.

Haughton, Dominique, Jonathan Haughton, Le Thi Thanh Loan and Nguyen Phong. *Shooting Stars and Sinking Stones – Wide jumps in Living Standards in Vietnam*.

Vu Duc Khanh, Vu Thu Thuy, Nguyen Thi Loan, Nguyen Quang Phuong and Le Hong Phong. *Labor and Employment through Two Vietnam Living Standards Surveys*.

Nguyen Thi Loan, Tran Thi Bich Van, Do Thi Phuong Lan, Vu Thanh Nam. *Health in Vietnam*.

Koch, Stefanie and Nguyen Bui Linh. *Child Malnutrition in Vietnam*

Le Thanh Tam, Dinh Lam Tan, Do Quy Toan, Nguyen Thi Viet Nga. *Informal Credit Market in Vietnam: Evidence of credit Rationing and Financial Mediation*.

Do Thien Kinh, Tran Quang, Bui Xuan Du, Le Do Mach, Nguyen Ngoc Mai, Lo Thi Duc. *Poverty in Vietnam – Evolution, Effects and Causes 1993-1998*.

Nguyen Quang Bang, Phi Mai Chi, Nguyen Dinh Chung, Le Hoang Dung, Vu Van Long, *Children in Poverty*.

Le Van Duy, Dominique Haughton, Jonathan Haughton, Do Anh Kiem, Le Dinh Ky. *Fertility in Vietnam: 1993-1998*.

Cao Nhu Nguyet, Nguyen Kim Dung, Bui Thai Quyen, Tran Bich Phuong. *Education in Poverty Reduction*.