

**CHINA LIVING STANDARDS SURVEY
1995 – 1997**

BASIC INFORMATION DOCUMENT

February 2003

ACRONYMS

CLSS	China Living Standards Survey
HRS	Household Responsibility System
PAF	Poverty Alleviation Fund
RCRE	Research Centre for Rural Economy

CONTENTS

1. OVERVIEW.....	1
2. SURVEY QUESTIONNAIRES.....	1
2.1. HOUSEHOLD QUESTIONNAIRE	1
2.2. COMMUNITY QUESTIONNAIRE.....	7
3. SAMPLE.....	10
4. ORGANIZATION OF THE SURVEY.....	11
5. USING THE DATA.....	12
5.1 DATA PROCESSING.....	12
5.2 LINKING PARTS OF THE HOUSEHOLD QUESTIONNAIRE	13
5.3 LINKING THE HOUSEHOLD TO THE PRICE & COMMUNITY DATA.....	14
6. CONSTRUCTED DATA SET.....	14
6.1 AGGREGATE EXPENDITURE, TOTEXP.....	14
APPENDIX A. HOW TO OBTAIN THE CLSS DATA.....	15
APPENDIX B. LIST OF VNLSS DATA SETS AVAILABLE.....	17
APPENDIX C. LIST OF RELATED DOCUMENTS	18
APPENDIX D. HOUSEHOLD EXPENDITURE CALCULATION	19

This document is part of an expanded program of documentation and further development of the Living Standards Measurement Study (LSMS) managed by Kinnon Scott in the Poverty Team of the Development Research Group (DECRG). It was prepared by Guo Li (EASRD), Paul Glewwe (consultant) and Diane Steele (DECRG).

China Living Standards Survey (CLSS), 1995-1997

Basic Information

1. Overview

China Living Standards Survey (CLSS) consists of one household survey and one community (village) survey, conducted in Hebei and Liaoning Provinces (northern and northeast China) in July 1995 and July 1997 respectively. Five villages from each three sample counties of each province were selected (six were selected in Liaoyang County of Liaoning Province because of administrative area change). About 880 farm households were selected from total thirty-one sample villages for the household survey. The same thirty-one villages formed the samples of community survey. This document provides information on the content of different questionnaires, the survey design and implementation, data processing activities, and the different available data sets.

2. Survey Questionnaires

2.1 Household Questionnaire

The household questionnaire contains sections that collect data on household demographic structure, education, housing conditions, land, agricultural management, household non-agricultural business, household expenditures, gifts, remittances and other income sources, and saving and loans.

For some sections (general household information, schooling, housing, gift-exchange, remittance, other income, and credit and savings) the individual designated by the household members as the household head provided responses. For some other sections (farm land, agricultural management, family-run non-farm business, and household consumption expenditure) a member identified as the 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 (employment), whenever possible, each member of the household was asked to respond for himself or herself, except that parents were allowed to respond for younger children. Therefore, in the case of the employment section it is possible that the information was not provided by the relevant person; variables in this section indicate when this is true.

The household questionnaire was completed in a one-time interview in the summer of 1995. The survey was designed so that more sensitive issues such as credit and savings were discussed near the end. The content of each section is briefly described below.

Section 0 SURVEY INFORMATION

This section mainly summarizes the results of the survey visits. The following information was entered into the computer: whether the survey and the data entry were completed, codes of supervisor's brief comments on interviewer, data entry operator, and related revising suggestion (e.g., 1. good, 2. revise at office, and 3. re-interview needed). Information about the date of interview, the names of interviewer, supervisor, data enterer, and detail notes of interviewer and supervisor were not entered into the computer.

Section 1 GENERAL HOUSEHOLD INFORMATION

1A HOUSEHOLD STRUCTURE

1B INFORMATION ABOUT THE HOUSEHOLD MEMBERS' PARENTS

1C INFORMATION ABOUT THE CHILDREN WHO ARE NOT LIVING IN HOME

Section 1A lists the personal id code, sex, relationship to the household head, ethnic group, type of resident permit (agricultural [*nongye*], non-agricultural [*fei nongye*], or no resident permit), date of birth, 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 to include "all the people who normally live and eat their meals together in this dwelling." Those who were absent more than nine of the last twelve months were excluded, except for the head of household. For individuals who are married and whose spouse resides in the household, the personal id number of the spouse is noted. By doing so, information on the spouse can be collected by appropriately merging information from the section 1A and other parts of the survey.

Section 1B collects information on the parents of all household members. For individuals whose parents reside in the household, parents' personal id 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 parent is alive, as well as their schooling and occupation.

Section 1C collects information for children of household members who are not living in home. Children who have died are not included. The information on the name, sex, types of resident permit, age, education level, education cost, reasons not living in home, current living place, and type of job of each such child is recorded.

Section 2 SCHOOLING

In Section 2, information about literacy and numeracy, school attendance, completion, and current enrollment for all household members of preschool age and older. The interpretation of pre-school age appears to have varied, with the result that while education information is available for some children of pre-school age, not all pre-school children were included in this section. But for ages 6 and above information is available for nearly all individuals, so in essence the data on schooling can be said to apply all persons 6 age and above. For those who were enrolled in school at the time of the survey, information was also collected on school

attendance, expenses, and scholarships. If applicable, information on serving as an apprentice, technical or professional training was also collected.

Section 3 EMPLOYMENT

3A GENERAL INFORMATION

3B MAJOR NON-FARM JOB IN 1994

3C THE SECOND NON-FARM JOB IN 1994

3D OTHER EMPLOYMENT ACTIVITIES IN 1994

3E SEARCHING FOR NON-FARM JOB

3F PROCESS FOR GETTING MAJOR NON-FARM JOB

3G CORVEE LABOR

All individuals age thirteen and above were asked to respond to the employment activity questions in Section 3. Section 3A collects general information on farm and non-farm employment, such as whether or not the household member worked on household own farm in 1994, when was the last year the member worked on own farm if he/she did not work in 1994, work days and hours during busy season, occupation and sector codes of the major, second, and third non-farm jobs, work days and total income of these non-farm jobs. There is a variable which indicates whether or not the individual responded for himself or herself.

Sections 3B and 3C collect detailed information on the major and the second non-farm job. Information includes number of months worked and which month in 1994 the member worked on these jobs, average works days (or hours) per month (per day), total number of years worked for these jobs by the end of 1994, different components of income, type of employment contracts. Information on employer's ownership type and location was also collected.

Section 3D collects information on average hours spent doing chores and housework at home every day during non-busy and busy season. The chores refer to cooking, laundry, cleaning, shopping, cutting woods, as well as small-scale farm yard animals raising, for example, pigs or chickens. Large-scale animal raising is excluded. Information about days spent helping others (including farming, building houses, not including "labor exchange" activities) without compensation in 1994 was also collected.

Information on searching and getting the major non-farm job was collected in Section 3E and 3F. Section 3E contains information on the member's estimated income which he/she could get from a non-farm job, reasons for giving up a non-farm job or not finding a non-farm job. Section 3F contains information about how the member got the job, what kind of help he/she received from other people, the relation between the member and the person who provided certain kinds help (e.g., information or recommendation for him/her to get the job).

Section 3G collects information of the household's corvee labor (*yiwu gong*). Corvee labor refers to a certain amount of work days farmers spend on collective projects (mainly collective infrastructure projects such as road construction, building and maintaining irrigation system) without compensation. Information includes the amount of corvee labor (number of

work days), whether the household fulfilled its “obligation”, amount of money the household paid for hiring other people to fulfill the corvee labor if the household did not fulfill it by itself.

Section 4 HOUSING

Section 4 contains basic information on housing for all households interviewed. Information was collected on the number of rooms in the dwelling, ownership status, construction materials of the house, the year the house was built, cost of construction, current value of the house, total construction area, living area, and the area used for production purpose. Area is measured in square meters.

Section 5 FARM LAND

5A GENERAL INFORMATION OF FARM LAND

5B PLOTS

5C PLOTS COMPARISON

Whenever is possible, a household member identified as the most knowledgeable on household’s farm land situation provided responses in Section 5. Section 5A records information on the village’s allocation of different tenure types land to the household since the Household Responsibility System (HRS) was introduced in early 1980s. Section 5A also collects information on land readjustment in past five years, land rental transaction among households, farmers’ willingness to engage in land rental activities, procurement quota, and other duties associated with land.

Section 5B contains basic information on each land plot currently farmed by the household. Information includes the time the household started to farm the plot, plot size, crops (summer and fall) planted, tenure types (e.g., responsibility land, ration land, private land), and some land quality indicators, such as irrigation condition, fertility category (high, medium, low, and very poor quality), topographic characteristics, and information on natural disaster on each plot in 1994 cropping year.

Section 5C is designed to compare the farmer’s production behavior on two plots with different land tenure types and associated land rights. Two plots, generally with different tenure types but farming the same crop, were selected for a careful investigation. Information includes labor input, animal traction, fertilizer use (organic fertilizer, phosphate fertilizer, and nitrogen fertilizer), other inputs (pesticide, herbicide), seed (conventional and hybrid), output, distance from plot to home, duties associated with each plot (procurement quota, taxes, and fee), village supplied unified production services if any, tenure type, some indicators of land rights (when the plot contract will expire, who makes the decisions about crop choice).

Section 6 AGRICULTURAL MANAGEMENT

6A ON-FARM LABOR

6B AGRICULTURAL INPUT

6C AGRICULTURAL ASSETS, FARM MACHINES AND EQUIPMENTS

6D1 CROP OUTPUT, DISPOSAL, STORAGE, AND MARKETING
6D2 CROP BY-PRODUCT
6E PROCESSING FARM PRODUCTS
6F LIVESTOCK PRODUCTION
6G VEGETABLE PRODUCTION, ORCHARD MANAGEMENT, FOREST AND
FISH POND ACTIVITIES

Whenever possible, a household member identified as the most knowledgeable on household's farming activities provided responses in Section 6. Section 6A collects information about on-farm labor activities, which refer to labor input in grain and cash crop production, such as basic farm land preparation, applying organic fertilizer, transplanting, irrigation activities, and harvesting. Labor input in small scale farm-yard animal raising also is included. On-farm labor, however, does not include labor input in commercialized large scale animal raising, vegetable production, orchard production, fish pond management, and corvee labor obligation.

Section 6B and 6C contain information about different agricultural inputs and agricultural assets (e.g., farm machines and equipment) respectively. Agricultural inputs refer to physical amount and/or money value of different kinds of fertilizer, herbicide and pesticide, electricity, diesel and gas. Expenses on hiring labor and labor exchange with other households also are included. In Section 6C, agricultural assets refer to tractors, mechanical plough, pump, mill or food processor, trailer for truck or tractor, and bullocks.

Section 6D1 collects information about crop output, disposal, storage, and marketing. Disposal includes self-consumption, use as seed and feed, gift, and being borrowed by other households. Marketing includes amount sold to state under procurement quota obligation and under negotiated price, amount and number of transactions at free market. Section 6D2 collects information on crop by-product.

Section 6E contains information about processing farm products, such as labor input (work days), total expense, frequency of transaction, and total income. Section 6F includes information on livestock production, such as type and number of livestock, marketing income, expenses for raising livestock etc.. Section 6G contains information on vegetable production, orchard management, forest and fish pond production activities, such as area, investment and its sources, total output, inputs (labor, seed, fertilizer), and taxes.

Section 7 FAMILY-RUN NON-FARM BUSINESS

7A GENERAL INFORMATION
7B ASSETS
7C DEBTS
7D EXPENDITURES
7E PROFITS

Section 7 gathers data on household non-farm businesses for the three most important enterprises operated by the household. The respondent is a household member identified as the most familiar with the business operation. Section 7A collects data on the ownership, type of

business, investment and its sources for each enterprises. Section 7B contains information on assets, such as buildings, vehicles, tools, and inventory of raw materials and products. Section 7C gathers information on enterprise debt and its structure (e.g., bank loan, loan from collective or cooperative foundation, and private loan). Expenditures over the last twelve months on wage, raw materials, taxes and other management activities are collected in Section 7D. In Section 7E, the respondent is asked to report total revenue, profits, and the amount (in money) of the enterprise's product was consumed by the household in the past twelve months.

Section 8 HOUSEHOLD CONSUMPTION EXPENDITURE

- 8A EXPENDITURES ON PURCHASING FOOD
- 8B HOME PRODUCED FOOD CONSUMPTION
- 8C NON-FOOD DAILY EXPENSES AND SERVICE EXPENDITURES
- 8D DURABLE GOODS EXPENDITURES

Section 8A collects detailed expenditure information on thirty-four items of market purchased food (including expenditure in restaurants) in past one year. Besides market purchases (including barter), Section 8B gathers information on consumption from home produced food (total thirty-two items) in past one year. In Section 8C, respondents were asked to recall the amount spent in past twelve months on daily expenses such as shoes, cloth, clothing, home repairs, public transport, paper supplies, kitchen equipment, medical services, jewelry, entertainment, cigarettes, soap. Amount, purchase price, year of purchase, and current value of durable goods owned were collected in Section 8D.

Section 9 GIFT-EXCHANGE, REMITTANCES, AND OTHER INCOME

- 9A Gifts and Remittances sent to others
- 9B Gifts and Remittances received from others
- 9C Other income

Section 9A records the information on amounts, relation and location of the recipients of remittances sent out from the household. Section 9B collects data on remittances received by the household. Section 9C contains information on other income (money and goods) from sources other than employment, such as subsidies from government, medicare subsidies, interest on saving, gifts, dowry or inheritance, and rent from land, equipment, and houses.

Section 10 CREDIT AND SAVINGS

- 10A MONEY AND GOODS BORROWED
- 10B MONEY AND GOODS LENT

Section 10A collects information on the amount of indebtedness of household members to people or institutions outside the household. General information includes number of times and number of different sources the household has ever borrowed money from in last five years. If money or goods have been borrowed, or borrowed and repaid by any household member in past five years, information is collected on those loans, including the source and amount of the

loan, interest, collateral requirement, repayment schedule, reason for each time of borrowing, amount paid in 1994, and amount still owed at the end of 1994. In Section 10B, the household is asked to report the number of times it lent money or goods to other people and institutions in past five years. If money or goods have been lent, or lent and repaid by other people and institutions in past five years, information is collected on those lending activities. The household is also asked to list different places (e.g., banks, credit union, loan to enterprises, leave it at home) to put away money which is not being used for a while, and to estimate maximum amount of money can be taken from own assets in facing of some kind of disaster or need to build a new house.

2.2 COMMUNITY QUESTIONNAIRE

A community survey was conducted in the same thirty-one sample villages from July to August in 1997. In each village, the enumerators interviewed three village leaders: the party secretary (*shuji*), the chairman of village committee or village leader (*cunzhuren* or *cunzhang*), and the village accountant (*kuaiji*). These three leaders generally are recognized as the most knowledgeable about village institutions now and in the past. The three leaders answered questions which included parts on off-farm labor, land management, local industrial management, local credit markets, periodic markets, agricultural input and output markets, and the local political institutions. Enumerators asked for information about two years – 1995 and 1988.

Enumerators also collected secondary data (from 1995, 1988, and 1980) on each village from statistical records of all villages. Information includes village basic information (e.g., total land area, total number of households, total population, income per capita, sown area, outputs and procurement quotas of different crops), and village budget information.

Section 1 VILLAGE BASIC INFORMATION AND LABOR MIGRATION

The Section 1 consists of two part: village basic information and labor migration. In village basic information part, information on size of village (number of households, number of groups, number of population and labor), infrastructure condition (e.g., road condition, number of bus or other passenger vehicle pass through the village, communication system, water system), village supplied agricultural production services (transportation of production inputs, land preparation service, irrigation service, crop protection service), allocation and adjustment of procurement quotas, taxes, and fees. Information on local farm labor market and farmers' corvee labor obligation was also collected.

Detailed information about out-migrants and in-migrants was collected in the labor migration part. Out-migrants are divided into five groups: permanent out-migrants, long-term out-migrants, commuters, self-employment, and villagers working in the village's collective and private enterprises. In-migrants are divided into three groups: permanent in-migrants, long-term in-migrants, and commuters. For each group of migrants, information on total number of migrants, education level, age structure, ownership type of employer, first three popular jobs, and moving destinations (for permanent, long-term, and commuters of out-migrants) was collected.

Section 2 CULTIVATED LAND SYSTEM

In Section 2, village leaders answered detailed questions about cultivated land management in the village. Information includes total cultivated land area and its structure in terms of different tenure types, such as private plots (*ziliu di*), rationed land (*kouliang tian*), responsibility land (*zeren tian*), and contract land (*chengbao tian*). For each land tenure type, land rights information on who hold the power on land readjustment, rental transaction, plot exchange, crop choice, input mix decision is collected. Information on obligation or duties associated with land, such as procurement quota, agricultural taxes, and local taxes, is also gathered. Specifically, village leaders reported the frequency, reasons, average magnitude, and timing of the village's land readjustment. Leaders also answered questions on other village-level regulations on land management, such as the village's rental activities and related institutions (including local rules concerning renting, number of households rented land in and out, and area of land engaged in rental transaction), process of village contracting out land, situation of land fragmentation, and village's attitude toward promoting large-scale farming.

Section 3 VILLAGE STORES, PERIODIC MARKETS, AND FARMERS COMMERCIAL ACTIVITIES

Section 3 has three parts. The first part gathers information on the stores in the villages, such as the number of department store, grocery store, restaurants and their ownership type (collective-run, private store). Information on where farmer buy their daily supplies and durable goods is also collected.

The second part of Section 3 collects detailed information on two periodic markets the villagers attend most frequently, including location of the market and distance from the village, how long the market has been established, the frequency of marketing operation, market size (in terms of estimated number of people attended the market in 1980, 88, and 96), the reason of the market size change. Information on farmers' major purposes for attending the market, places the most buyers and sellers come from, whether the market is a specialized market is also collected in the second part of the Section.

The third part of Section 3 contains information on farmers' selling (proportion by different buyers), exchange, and consumption of major agricultural products, such as major grain crops, cash crop, and major livestock product. Information on farmers purchase (proportion by difference sources) of staple food and non-staple food is also collected. Price difference between market and state grain station of paddy, rice, wheat and flour in 1988 and 1996 is also collected.

Section 4 VILLAGE-RUN ENTERPRISE

Section 4 records detailed information on one village-run enterprise. If the village has more than one enterprise, the enumerator selected one of them randomly. Information about the enterprise includes the year the business was established, major products (codes), basic information on the manager (gender, age, education, the place he/she came from, the number of years he/she has

worked in the enterprise, personal experience - such as a veteran, or was a manager of a state-owned enterprise, or was a government official), the relation between manager and village cadres, manager's decision power on business affairs, the type of contract signed between village and manager etc. Information on the size of the enterprise in terms of number of workers, amount of investment, value of fixed asset, and profit is also recorded. Finally, Section 4 collects information on major methods the enterprise used to purchase raw materials and sell its products.

Section 5 RURAL CREDIT MARKET

Section 5 includes basic information on farmers' saving, borrowing (from both financial institution and private person), and allocation of poverty alleviation fund (PAF) if applicable. Village leaders estimated the proportion of farm households which have deposits in the bank and the average amount of saving these households have. Information on major financial institutions, such as distance from the village, the transportation time needed to go these institutions, market share of different institutions, and degree of convenience to withdraw money, is also collected. Section 5 lists out some typical activities (e.g., buying production inputs, buying draft animals, running a small business, constructing a house, dealing with illness) and asks village leaders to estimate the proportion of farm households that need to borrow money, the average amount they need to borrow, and the three major borrowing sources. Village leaders also estimated the proportion of private borrowing out of total numbers of borrowing, and the interest rate of private borrowing if applicable. Finally, village leaders reported basic information on the allocation of poverty alleviation fund (PAF), such as sources of PAF, interest rate, the average income level of PAF beneficiaries, and major activities the PAF were invested.

Section 6 SEED MARKET

Section 6 collects information on local seeds market for rice, wheat, corn, and vegetable. Information in this section includes proportion of regular seeds farmers purchased from market, the price of regular and hybrid seeds, market share of different seeds supplying sources (includes private traders).

The data from Section 6 are not available for distribution.

Section 7 CORN PRODUCTION TECHNOLOGY

Hebei and Liaoning are two major corn production provinces in China. Section 7 contains basic information such as major uses of corn (e.g., for staple food, feed, brewing liquor), the history of using hybrid corn (which year started, the area sown), and three major constraints of increasing the corn yield.

Section 8 CHEMICAL FERTILIZER MARKET

Section 8 collects information on the supply and its changes of subsidized (*pingjia*) and non-subsidized (*yijia*) major kinds chemical fertilizers (nitrogen, phosphate, and potash). Information includes whether there is subsidized fertilizer, how it is allocated to farm households,

and the proportion of subsidized fertilizer accounts of total amount fertilizer used by farmers. Enumerators also asked village leaders to recall the information pertaining to the development of fertilizer markets, for example, the changes to the subsidized fertilizer supply in past eight years and the reasons for these changes (e.g., when and why state stopped or resumed supplying subsidized fertilizer), whether or not the state commercial units guaranteed the supply of subsidized fertilizer, the major sources from which farmers buy chemical fertilizers, and if there was a serious shortage of fertilizer in past eight years.

Section 9 VILLAGE LEADERSHIP

Section 9 contains basic information on village-level political institutions. Information in this section includes when the Village Committee (*cunweihui*) was established, the frequency of election of members to the Village Committee, village leaders' role in the village-run enterprises, and the election of party secretary. Enumerators also collected information on formation, frequency of meeting, and general attendance of the Village Assembly, the Village Representative Conference, and the Village Enterprises Management Committee. The section also records information about party secretary (*shuji*) and chairman of the Village Committee (*cunzhuren*), such as age, education level, years as a village leader, veteran or not, the job before being elected as a leader. Finally, the section collects brief information on village's most recent and first time election, such as the year of election, content of the election, and whether or not the election is competitive (i.e., the number of candidates are more than number of positions).

Section 10 FORESTRY

Section 10 on Forestry was not administered in the 1997 survey. It was used in an earlier administration of the survey. Section 10 is only included in the Chinese version of the Community Questionnaire.

3. Sample

The CLSS sample is *not* a rigorous random sample drawn from a well-defined population. Instead it is only a rough approximation of the rural population in Hebei and Liaoning provinces in Northeastern China. The reason for this is that part of the motivation for the survey was to compare the current conditions with conditions that existed in Hebei and Liaoning in the 1930's. Because of this, three counties in Hebei and three counties in Liaoning were selected as "primary sampling units" because data had been collected from those six counties by the Japanese occupation government in the 1930's. Within each of these six counties (*xian*) five villages (*cun*) were selected, for an overall total of 30 villages (in fact, an administrative change in one village led to 31 villages being selected). In each county a "main village" was selected that was in fact a village that had been surveyed in the 1930s. Because of the interest in these villages 50 households were selected from each of these six villages (one for each of the six counties). In addition, four other villages were selected in each county. These other villages were not drawn randomly but were selected so as to "represent" variation within the county. Within each of these villages 20 households were selected for interviews. Thus the intended sample size was 780 households, 130 from each county.

Unlike county and village selection, the selection of households within each village was done according to standard sample selection procedures. In each village, a list of all households in the village was obtained from village leaders. An “interval” was calculated as the number of the households in the village divided by the number of households desired for the sample (50 for main villages and 20 for other villages). For the list of households, a random number was drawn between 1 and the interval number. This was used as a starting point. The interval was then added to this number to get a second number, then the interval was added to this second number to get a third number, and so on. The set of numbers produced were the numbers used to select the households, in terms of their order on the list.

In fact, the number of households in the sample is 785, as opposed to 780. Most of this difference is due to a village in which 24 households were interviewed, as opposed to the goal of 20 households

4. Organization of Survey

The data were collected as a joint effort of China’s Research Centre for Rural Economy (RCRE), which is part of China’s Ministry of Agriculture, and a North American team consisting of Loren Brandt (University of Toronto), Paul Glewwe (World Bank), and Scott Rozelle (Stanford University and University of California at Davis). The RCRE team was led by Mr. Bai Nan-Shang. The household survey data were collected in the summer of 1995. The interviewers consisted of students in the Department of Sociology at Beijing University. The interviewers were trained over a period of two weeks immediately before the beginning of the survey. They were divided into teams, each of which had interviewers and a data entry operator. Each team covered two counties, one each in Hebei and Liaoning. The three North American collaborators and Mr. Bai served as supervisors. Each team worked in a county for 3-4 weeks.

Data entry was done in the villages where the interviews took place, either on the day of the interview or on the following day. A customized data entry program was used to detect errors and inconsistencies, which allowed the interviewers to return to the households to correct any errors found. Data collection began in July 1995 and ended in August 1995. No major problems occurred in the field work, and the data appear to be of very good quality.

Although village level data were also collected in the summer of 1995, analysis of those data revealed many problems and all 31 villages were revisited in July and August of 1997 to collect village level data. In each village, the enumerators interviewed three village leaders: the party secretary (*shuji*), the chairman of village committee or village leader (*cunzhuren* or *cunzhang*), and the village accountant (*kuaiji*). These three leaders generally are recognized as the most knowledgeable about village institutions now and in the past.

5. Using the Data

It is strongly recommended that the data be used *with* the questionnaires. English translations are available for both the household and the village questionnaires, as described in Appendix C. The questionnaires contain the exact wording of the questions and interviewer instructions. Since the questionnaires used to collect the data were in Chinese, it is also advisable to check the Chinese versions of the questionnaires in case of doubt or confusion regarding a question.¹ All codes, except those for occupation, economic sector, province, and crops and fruits, are contained in the questionnaire on the same page as the questions. The codes that are not on the same pages as the question in the questionnaire are listed in the back of the English questionnaire.

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 not relevant 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 “>>” followed by a number (e.g. >>4) after a particular response indicates which question should be asked if that reply is given. This implies skipping over other questions. Copies of both the household and the community questionnaires (English and Chinese) are available from the World Bank (see Appendix C).

The household and community data are available free of charge through the World Bank web site. Information on how to download the data is provided in Appendix A. A complete list of data sets is provided in Appendix B. The data sets are available in SAS Portable, STATA, and ASCII formats.

5.1 Data Processing

This section describes the different data processing stages between the actual interviewing and the final datasets 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 dataset stage.

¹ The Chinese version of the Community Questionnaire is only available through the LSMS Office. To obtain a copy, see Appendix C.

All responses obtained from the household interviews were recorded in the household questionnaires. These were then entered into the computer, in the field, using data entry programs written in BASIC. The data produced by the data entry program were in the form of household files, i.e. one data file for all of the data in one household/community questionnaire. Thus for the household there were about 880 data files.

These data files were processed at the University of Toronto and the World Bank to produce datasets in statistical software formats, 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 66 data files were created. Members of the survey team checked and corrected data by checking the questionnaires for original recorded information. We would like to emphasize that correction here refers to checking questionnaires, in case of errors in skip patterns, incorrect values, or outlying values, and changing values if and only if data in the computer were different from those in the questionnaires. The personnel in charge of data preparation were given specific instructions not to change data even if values in the questionnaires were clearly incorrect. We have no reason to believe that these instructions were not followed, and every reason to believe that the data resulting from these checks and corrections are accurate and of the highest quality possible.

The screen files were then brought to World Bank headquarters in Washington, D.C. and uploaded to a mainframe computer, where they were converted to "standard" LSMS formats by merging datasets to produce separate datasets for each section with variable names corresponding to the questionnaires. In some cases this has meant a single dataset for a section, while in others it has meant retaining "screen" datasets with just the variable names changed.

5.2 Linking Parts of the Household Survey

Each household has a unique identification number which is contained in the variable HID. Values for this variable range from 10101 to 60520. The first number is the code for the six counties in which data were collected, the second and third digits are for the villages within each county. Finally, the last two digits of HID contain the household number within the village.

Data for households from different parts of the survey can be merged by using the HID variable which appears in each dataset of the household survey. To link information for an individual use should be made of both the household identification number, HID, and the person identification number, PID.

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.

5.3 Linking the Household to the Community Data

The community data have a somewhat different set of identifying variables than the household data. Each community dataset has four identifying variables: province (code 7 for Hebei and code 8 for Liaoning); county (six two digit codes, of which the first digit represents province and the second digit represents the three counties in each province); township (3 digit code, first digit is county, second digit is county and third digit is township); and village (4 digit code, first digit is county, second digit is county, third digit is township, and third fourth digit is village).

6. Constructed Data Set

Researchers at the World Bank and the University of Toronto 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 set has been created using only the original (raw) data sets. A researcher could construct a somewhat different data set by incorporating different assumptions.

6.1 Aggregate Expenditure, TOTEXP

The dataset TOTEXP contains variables for total household annual expenditures (for the year 1994) 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 algorithm used to calculate household expenditures are detailed in Appendix D. The dataset also contains the variable HID, which can be used to match this dataset to the household level data set. Note that all of the expenditure variables are totals for the household. That is, they are *not in per capita terms*. Researchers will have to divide these variables by household size to get per capita numbers. The household size variable is included in the data set.

Appendix A

How to Obtain the CLSS Data

The CLSS data are publicly available. Copies of the documentation (Basic Information Document and questionnaires) and data can be downloaded free of charge from the LSMS Web Site:

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

or obtained by mail through the LSMS Office (see address below). It is recommended that individuals who are interested in using the data for analyses read the documentation prior to requesting copies of the data to ensure that the data will be useful to their analyses.

There is an open access policy for the CLSS data. This means that potential users do not need to obtain the permission of the RCRE to receive a copy of the data. Users can request copies of the data and documentation from:

LSMS Database Manager
Poverty and Human Resources
Development Research Group
The World Bank
1818 H Street, NW
MSN MC3-306
Washington, DC 20433
tel: (202) 473-9041
fax: (202) 522-1153
e-mail: lsms@worldbank.org

The following information should be included in the request: (a) a brief description of the research that will be done with the data; (b) an indication of the format in which the user prefers to receive the data (ASCII, SAS Portable, or STATA); and (c) if requested from the World Bank, a check made out to the World Bank for the processing fee. Data and documentation will be provided on CD in the user's requested format.

The World Bank charges a processing fee charged for providing the data through the LSMS Office. For the most current information on the processing fee, contact the LSMS office at the above address, send an e-mail to lsms@worldbank.org or check the web site.

Individuals who receive copies of the data agree to: (a) cite the Research Centre for Rural Economy as the collector of the data in all reports, publications and presentations; (b) provide copies of all reports, publications and presentations to the Research Centre for Rural Economy and the Poverty and Human Resources Division of the Development Research Group of the World Bank (see address above); and (c) not pass the data to any third parties for any reasons.

Researchers found to be in violation of these agreements will not be able to receive copies of other data sets from the LSMS Office in the future.

The researcher should further note that once received, the data cannot be passed on to a third party for any reason. Researchers are requested to provide copies of their results to the World Bank.

Appendix B

List of CLSS Data Sets Available

The following data sets are available on 3 1/2" diskette. All are available SAS Portable, STATA and ASCII files.

	CLSS DATA SETS
HOUSEHOLD (numbers correspond to sections of the household survey)	S00
	S01A1, S01A2, S01B, S01C1, S01C2
	S02
	S03A, S03B, S03C, S03D, S03E, S03F, S03G
	S04
	S05A1, S05A2, S05A3, S05A4, S05A5, S05A6, S05B, S05C1, S05C2, S05C3, S05C4
	S06A1, S06A2, S06B1, S06B2, S06B3, S06B4, S06B5, S06C, S06D1, S06D2, S06D3, S06D4, S06E1, S06E2, S06F1, S06F2, S06F2N, S06G1, S06G2
	S07A1, S07A2, S07B, S07C, S07D, S07E
	S08A1, S08A2, S08B1, S08B2, S08C, S08D1, S08D2
	S09A1, S09A2, S09B1, S09B2, S09C
S10A1, S10A2, S10B1, S10B2	
VILLAGE	NPT0101, NPT0102, PT0103, PT0104, NPT0105, NPT0106, NPT0107, NPT0108A, NPT0108B, PT0109, PT0110, PT0111, NPT0112A, PT0112B, PT0201A, PT0201B, PT0201C, PT0202A, PT0202B, PT0203A, PT0203B, PT0204A, PT0204B, PT0205, NPT0301, NPT0302, PT0303, PT0304, PT0305, NPT0401, PT0402, PT0403, PT0404A, PT0404B, PT0405A, PT0405B, PT0501A, PT0501B, PT0502A, PT0502B, PT0701, PT081A, PT0801B, PT0802A, PT0802B, PT0803, PT0803B, PT0804, PT0901A, PT0901B, PT902A, NPT0902B.
Constructed	TOTEXP

Appendix C

List of Related Documents

The following documents can be obtained from the World Bank LSMS Office free of charge. See Appendix A for information on how to contact the LSMS Office.

A. Questionnaires

1. Household Questionnaire (English & Chinese versions)
2. Community Questionnaire (English & Chinese versions)

B. Basic Information Document (English)

C. Other

1. Household File Codebook
2. Village File Codebook

Appendix D: Household Expenditure Calculation

To get a measure of household consumption expenditures, the CLSS tried to collect information, as accurately as possible, on all expenditures that enhance household welfare in the year preceding the interview. The CLSS contains expenditure information on various food and non-food items. The way in which this information is used to calculate a measure of household expenditure in the year preceding the interview is described in this appendix.

1. Food expenditures

Information on household expenditure on food items is available in Sections 8A (expenditures on purchasing food, including barter) and 8B (home produced food consumption). In Section 8A, for 25 food items/categories information is available on the quantity and value of market purchases and the quantity bartered. For other 8 food items/categories (including eat in restaurants) information is available on either the value of market purchases or the quantity bartered or both. Experience of pilot survey for designing questionnaire indicated that farmers had some difficulties to recall accurate quantity they purchased for these 8 food items, and this is the reason why CLSS did not collect information on purchased quantity of them. In Section 8B, for 26 food items/categories information is available on the consumed quantity of home produced food. Information on cigarette and tobacco is collected in Sections 8A and 8B, however, we put expenditures on cigarette and tobacco into category of non-food expenditure.

To get a total value of household food expenditure, we need to find an appropriate price for each food item to convert quantities into values, for example, home produced consumption, bartered amount, and purchased amount (if household did not report values). The procedures to get appropriate prices, complete expenditure of market purchased food, and self-produced amount are explained below.

1.1 Getting the Prices

For those 25 food items/categories information on both purchased quantity and value are available, we first divided value (s08a1b) by the quantity (s08a1a) to get the household level price. To get a representative price of each food item by village, and also for getting rid of impacts of extreme cases, we pick the median price of each food item as the representative price.² In case this village-level price is missing, we use county median price as a substitute, and if the county-level price is still missing, we use sample median price as a substitute. For detail on price calculation, see program price.do.

For other 8 food items, we could not get prices from the survey since CLSS did not collect information on purchased amount for reasons explained above. To convert bartered

² Generally, average price is more sensitive to the extreme cases than median price. For example, in case of vegetable (code 124), the average price of vegetable is 3.4 yuan/kg, and the median is 1.0 yuan/kg, and the latter is more reasonable. Therefore, we decide to use median price eventually.

and/or self-produced amounts of these foods into value, we use 1994 prices from China Price Statistical Yearbook (1995). These prices are listed at the end of the appendix.

1.2 Expenditure on Market Purchasing Food

Section 8A contains information on market purchased food. To get the complete expenditures on market purchased food, we first convert purchased amount into values by using the above prices (if the household did not reported value of these market purchased value). Then we convert bartered amount into the values. Finally, we sum the above two calculated values and household reported values (s08a1b) together to get the complete expenditures on market purchased food (expenditures on cigarette and tobacco are excluded for they are counted as non-food expenditure).

1.3 Value of Home Produced Consumption

Section 8B contains information on home produced food consumption. To get the complete value of home produced consumption, we convert all quantities into the value by using the prices get above. Then we add all values together to get the total value of home produced food consumption (expenditures on cigarette and tobacco are excluded).

2. Nonfood Expenditure

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

2.1 Daily Expenses and Service Expenditure

Section 8C contains information on 29 purchased nonfood items. These includes expenditures on cigarettes, tobacco, cooking fuel, soap and detergents, haircut, cosmetics, entertainment, electricity, jewelry, books (not for schooling), medical services, etc.. Total expenditure on these items is obtained by adding them together. 12 items (item codes in parentheses) are excluded from household expenditures for several reasons. Expenses on lotteries tickets (337) 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. The following items in this section are also not included because they do not increase economic well-being: insurance (333), education foundation (335), donation (336), fines (339) and gift, transfer (340). The following items are not included because expenditures on them are likely to be reflected in the current value of the relevant consumer durable or non-durable: hardware (318), bicycle tire and parts (319), maintenance expense for all transportation products (320), construction materials (322), home improvement, paint (323), grain processing (338).

2.2 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 living. Excluding these 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. Two items of durable goods (codes in parentheses), three-wheel cart (415) and hand pushing cart (416), are excluded from the household expenditures because are mainly used for productive purposes.

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, current value at the time of interview, and the 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:

$$d = 1 - (VT / VB)^{1/t}$$

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 (note: we actually used average depreciation rate, we can change it back to median). These depreciation rates were calculated using “overall industrial products rural retail price indices” from 1978-1995 (China Statistical Yearbook, 1996). By doing so, we assume that prices before 1978 are same, which is an acceptable assumption given the there was almost no inflation in China's central planned economy before 1978.

Generally, the second step should consist 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. However, in 1994, the real interest

rate was negative because the inflation rate was higher than interest rates.³ Therefore, we only use depreciation rate to calculate use value:

$$UseValue = VT(1 + d)^{1.5} \cdot d$$

where $VT(1 + d)^{1.5}$ is the value of the consumer durable at the beginning of 1994 (because we did our survey in mid 1995, so power 1.5 is used). For detail about calculation, see program durable.do.

2.3 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 rural China is not possible to use this procedure because there is no competitive rental market for housing. Several rounds pilot survey for preparing questionnaire for CLSS showed that there is almost no households live in dwellings they do not own. Therefore, in housing section of formal questionnaire, there is even no questions about house renting activities.

To get the house depreciation rate and eventually to obtain the “use value” of dwellings, the following method is used. Assuming the current value of the house is determined by some house characteristics, such as construction materials, the number of years the house has been built, the number of rooms, the size of living area etc., we run the following regression:

$$\log(Y) = \sum_{i=1}^5 \alpha_i D_i + \beta N + \delta A + \rho T + \sum_{j=1}^{30} \eta_j V_j + \varepsilon$$

where Y is the current value of house, D 's are five dummies for characteristics of house (Shared with other households? Construction material is straw, mud, bricks? Is a building more than one story?), N is number of rooms, A is total area of rooms, T is the number of years the household owned the house, V 's are thirty village dummies, and ε is error term.

Coefficient of T , ρ , is the depreciation rate of house because

³ According the China Statistical Yearbook (1996), the inflation rate was 24.1 percent in 1994, and the interest rate for one year fixed deposits was 10.98 percent.

$$\rho = \frac{1}{Y} \cdot \frac{\partial Y}{\partial T}$$

The value of ρ shows that the average depreciation rate of house is about 5.83, which is regarded as a reasonable value.

After getting the depreciation rate of house, a similar approach as in the case of durable goods is used to calculate the “use value” of dwellings. For the same reason as in the case of durable goods, we also ignore the effects of the negative real interest rate. Instead of applying ρ to the reported value of a dwelling, the predicted housing value from above regression is used, since reported values are likely to have substantial random variation related to respondent knowledge, etc. Therefore, the “use value” of dwellings is calculated by following formula:

$$UseValue = e^{\bar{Y}} \cdot (1 + \rho)^{1.5} \cdot \rho.$$

where e is the base of natural logarithm, \bar{Y} is the predicted value of house (in log form), and $e^{\bar{Y}} \cdot (1 + \rho)^{1.5}$ is the value of house at the beginning of 1994 (because we did our survey in mid 1995, so power 1.5 is used). For detail about calculating use value of house, see program house.do.

2.4 Education

For each household member, the total expenditures on tuition, registration, uniforms, books, etc. are obtained in Section 2 (variable s0207). If the household member received fellowship or scholarship in 1994, the information on the amount of fellowship or scholarship is also obtained in Section 2 (variable s0209).

2.5 In-kind Wages

Wage received in-kind are a form of household consumption not measured elsewhere, and therefore should be included in household expenditures. Sections 3B and 3C, contains information on in-kind wages from employers of major non-farm job and secondary non-farm job in 1994 (variables s03b18 and s03c18). In-kind wages refer to the benefits employee received in form of goods instead of cash, such as medicine and other health-related supplies, food, beverages, and other gifts given during holidays.

3. Data problems and Corrections

In theory, a variable-recall procedure, such as the one used in the CLSS, 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. A variable-recall procedure, however, 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:

(1) Unit prices were calculated by dividing expenditure values (for food items) by quantities, and their distributions 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. To make sure the outlying values are not results of entry errors, we checked original forms. Some prices are still extremely high or low after the checking, the criteria for dealing with those cases is that if any reasonable change will only affect less than 5 percent (on average) of total expenditure, we will not make any changes. All corrections are noted in the Sate program available from the World Bank. Note that no corrections have been made to the original data.

(2) Shares of food items, “use value” of durable goods and house in total annual expenditures were checked to identify outliers. Note that these patterns were examined only to identify extremely inconsistent observations and then check the component parts of the expenditure variable for potential data entry errors. One specific case in food consumption is noted. Household 40410 had extremely high value of market purchased foods. By checking the components of market purchased food items, it is found that eating in restaurants accounted 70 percent of total market purchased food (50,000 out of 71,000 yuan). Because the total expenditures of the household belong top 5 percent group, which implies that the household is quite well-off, we made no changes on the household’s food consumption. Average share of durable goods “use value” in total expenditure (without durable goods use value) is 4.6 percent. But in six household’s shares are among 30-50 percent of total expenditure. After checking with original forms, we found these relatively high proportions are generally reasonable. For example, among these six households, 2 reported had car, 3 reported had motor vehicles, 4 reported new refrigerators. In house value case, the average share of house “use value” in total expenditure (without house use value) is 9.89 percent. There are four observations range from 50-69 percent which still can be taken as reasonable.