China - Living Standards Survey 1995 -1997

Research Centre for Rural Economy, The World Bank

Report generated on: October 26, 2020

Overview

Identification

ID NUMBER
CHN_1995-1997_LSS_v01_EN_M_v01_A_OCS

Overview

ABSTRACT
China Living Standards Survey (LSS) 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.

KIND OF DATA
Sample survey data [ssd]

UNITS OF ANALYSIS
Households

Scope

NOTES
The scope of the study included:

(a) HOUSEHOLD QUESTIONNAIRE

Section 1: General Household Information

Section 2: Schooling

Section 3: Employment

Section 4: Housing

Section 5: Farmland

Section 6: Agricultural Management

Section 7: Family-Run Non-Farm

Section 8: Household Consumption

Section 9: Gift-Exchange, Remittances, and Other Income

Section 10: Credit and Savings

(b) COMMUNITY QUESTIONNAIRE

Section 1: Village Basic Information and Labour Migration

Section 2: Cultivated Land System

Section 3: Village Stores, Periodic Markets, and Farmers Commercial Activities

Section 4: Village-Run Enterprise

Section 5: Rural Credit Market

Section 6: Chemical Fertilizer Market

Section 7: Village Leadership

TOPICS

Topic	Vocabulary	URI
Agriculture & Rural Development	FAO	
Food (production, crisis)	FAO	
Financial Sector	FAO	
Migration & Remittances	FAO	
Infrastructure	FAO	

Coverage

GEOGRAPHIC COVERAGE
Regional

Producers and Sponsors

PRIMARY INVESTIGATOR(S)

Name	Affiliation
Research Centre for Rural Economy	
The World Bank	

OTHER PRODUCER(S)

Name	Affiliation	Role
The World Bank		Technical assistance

Metadata Production

METADATA PRODUCED BY

Name	Abbreviation	Affiliation	Role
Office of Chief Statistician	OCS	Food and Agriculture Organization	Adoption of metadata for FAM
Development Economics Data Group	DECDG	The World Bank	Generation of the DDI

DDI DOCUMENT VERSION
CHN_1995-1997_LSS_v01_EN_M_v01_A_OCS_v01

DDI DOCUMENT ID
DDI_CHN_1995-1997_LSS_v01_EN_M_v01_A_OCS_FAO

Sampling

Sampling Procedure

The China LSS 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 North-eastern 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

Questionnaires

No content available

Data Collection

Data Collection Dates

Start	End	Cycle
1995-07	1995-08	N/A
1997-07	1997-08	N/A

Data Collection Mode

Face-to-face [f2f]

Data Processing

Data Editing

(a) DATA ENTRY
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.

(b) DATA EDITING
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. 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. 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). 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. 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.

Data Appraisal

No content available