

Individual Income Variable Construction

Individual income is conceptualized as the sum of all sources of income and revenue minus expenditures for one household member. It is not a simple division of household income evenly among household members (that is per capita income, which is computed as part of household income). Rather, individual income is built by adding each person's income source.

There are questions about seven potential sources of income in the questionnaires: business, farming, fishing, gardening, livestock, non-retirement wages, and retirement income. While household income includes the income from subsidies and other income, these cannot be allocated to individuals in the household and are not considered part of individual income. Examples of income sources **not** included are subsidies for health, one-child, food, utilities, etc.; gifts; rent; and in-kind payments other than in the context of one of the income sources above. Details on each source follow.

Data come from the longitudinal files for each income source. These files have all of the data for both an individual and for an individual at each wave. Not all individuals or individuals are represented in each file. In general, if an individual was present in one of these files, it was assumed that the individual reported income from that source. Similarly, if an individual was not represented in the file, it was assumed that the individual did not report any income from that source. More recent waves included filter questions to help determine when an individual had income from each source, and these filters were used when possible.

When an individual was determined to have income from a source (by filter question or by presence in the longitudinal file), but the data were incomplete for that individual, an attempt was made to impute the missing data. In order of preference, imputation was usually based on the individual's previous and subsequent waves, the mean of individuals in the community, or the mean in the city/county. (Not all income sources were handled this way - see details below under each source.) If fewer than three individuals supplied data at any of these levels, imputation was not done at that level. If a value was imputed for a respondent, an impute flag was given the value of 1. This allows you to drop respondents with imputed data if you so choose.

After calculating individual income from each source, total individual income was constructed as the sum from all seven sources. This variable is called *indinc*. The value at each wave was then inflated to 2006 Yuan currency values. This variable is called *indinc_cpi*. The methods are described on the following pages.

Note that many households reported negative net income from business, farming, gardening, fishing, or raising livestock, which causes individual income for that activity to be negative as well. This is especially apparent for livestock: looked at over time, 61% of households who raised livestock reported higher expenses than revenues during at least one of the waves of data collection. This probably is due to the cyclic nature of raising livestock, but other activities like farming and gardening are subject to annual weather

differences, and market prices can vary from year to year. As the project director noted, "It's normal that a person invested big money in livestock or other businesses and earned nothing in one year, might gain or lose a lot in the second year, and did nothing in the third year. I heard such stories several times when I participated in data collection or supervised the fieldwork."

Business Income

Variable: INDBUS—Total individual net income from all businesses operated by household that the individual participated in. Component of INDINC, total individual income.

Data Files: H07BUSI- individual level, time spent working in HH business
M07BUSN-household level, net income from HH business

Source: H2, Business type
H3, Revenue from this business
H4, Expenses
H6, Months worked in HH business last year
H7, Days per week worked in HH business last year
H8, Hours per day worked in HH business last year

Basic Algorithm: Individual proportion of net HH income from household businesses. The proportion is based on reported hours spent working in each household business. Calculations are done by business type, i.e., commerce, service, manufacturing, peddler, construction, and other. HH income for each type is summed for each HH within type. Then the individual proportion is calculated for each business type as the time each HH member spends working in that business type divided by the sum of the time all HH members spend working in that business type. The net income for each business type is then apportioned to each individual who reported working in the business type. For 1989, variables H6-8 are not available, so income is distributed equally among members who report the same HH business type in H2.

Imputation of missing values

Logic: Presence of a business type record for household is held as proof family had an individual business. Similarly, the presence of a business type record in the individual data means the individual worked in that family business. Any missing revenue or expense is imputed if source of imputation is available.

Impute missing H6-8 in the following order of preference:

1. household average for that business type
2. community average for all business types
3. county average for all business types
4. distribute income equally among people in household who reported working on each type of HH business

Farming Income

Variable: INDFARM - Total individual net income from farming (see separate Gardening macro). Component of INDINC, total individual income.

Data Files: H07FARMG-Individual time spent farming
FARMyyyy-HH net income from farming

Source: E4A, months farmed last year
E4B, days farmed per week last year
E4C, hours farmed per day last year
E2A, worked in HH farm/orchard last year (from 2004 on)
E4, 12-month average hours farmed per week (1989 only)

Basic algorithm: Individual proportion of net HH income (HHFARM) from household farming, where the proportion is based on the reported hours spent farming. The individual proportion is calculated as the time each HH member spends farming divided by the sum of the time all HH members spend farming. Convert E4A-C to 12-month average hours worked per week to match 1989 data where that's the only variable available.

Imputation of Missing Values

If rural, and filter variables indicate farming activity, missing data is imputed where available. Because weather has such a great impact on farming activities, it was determined that previous and subsequent waves should not be used to impute farming variables. The Household mean is used first if at least one other HH member reported farming data for that variable. If the value was still missing, the Community or County mean values were used if at least 3 other individuals in the community or city/county reported farming data for that variable. If the value is still missing, all farming is allocated equally to the HH members who farm.

Fishing Income

Variable: INDFISH - Individual income from fishing. Component of INDINC, total individual income.

Data Files: H07FISHI - Individual time spent fishing for HH fishing business
FISHyyyy - Household net income from fishing

Source: G4A, months fished last year
G4B, days fished per week last year
G4C, hours fished per day last year
G2, filter: worked in fishing last year (from 2004 on)
G4, 12-month average hours fished per week (1989 only)

Basic Algorithm: Individual proportion of net HH income (HHFISH) from household fishing, where the proportion is based on the reported hours spent fishing. The individual proportion is calculated as the time each HH member spends fishing divided by the sum of the time all HH members spend fishing. Convert G4A-C to 12-month average hours worked per week to match 1989 data where that's the only variable available.

Imputation: If filter variables indicate fishing activity, missing data is imputed where available. The Household mean is used first if at least one other HH member reported fishing data for that variable. If the value was still missing, the Community or County mean values were used if at least 3 other individuals in the community or city/county reported fishing data for that variable. If the value is still missing, all fishing is allocated equally to men aged 21-60 in the HH.

Gardening Income

Variable: INDGARD - Total individual net income from gardening (see separate Farming macro). Component of INDINC, total individual income.

Data Files: H07FARMG - Individual time spent gardening
GARDyyyy - HH net income from gardening

Source: D3A, months gardened last year
D3B, days gardened per week last year
D3C, hours gardened per day last year
D2A, worked in HH garden last year (from 2004 on)
D3, 12-month average hours gardened per week (1989 only)

Basic algorithm: Individual proportion of net HH income (HHGARD) from household gardening, where the proportion is based on the reported hours spent gardening. The individual proportion is calculated as the time each HH member spends gardening divided by the sum of the time all HH members spend gardening. Convert D3A-C to 12-month average hours worked per week to match 1989 data where that's the only variable available.

Imputation of Missing Values

If any of D3A-C is not missing, or if filter variable indicates farming activity, missing data is imputed where available. The Household mean is used first if at least one other HH member reported farming data for that variable. If the value was still missing, the Community or County mean values were used if at least 3 other individuals in the community or city/county reported farming data for that variable. If the value is still missing, all gardening is allocated equally to women aged 21-60 in the HH (this is rare).

Livestock Income

Variable: INDLVST - Total individual net income from raising livestock. Component of INDINC, total individual income.

Data Files: H07LIVEI - Individual time spent raising livestock
GARDyyyy - HH net income from raising livestock

Source: F4A, months raised livestock last year
F4B, days raised livestock per week last year
F4C, hours raised livestock per day last year
F2A, raised livestock last year (from 2004 on)
F4, 12-month average hours raised livestock per week (1989 only)

Basic algorithm: Individual proportion of net HH income (HHLVST) from household livestock business, where the proportion is based on the reported hours spent raising livestock. The individual proportion is calculated as the time each HH member spends raising livestock divided by the sum of the time all HH members spend raising livestock. Convert F4A-C to 12-month average hours worked per week to match 1989 data where that's the only variable available.

Imputation of Missing Values

If any of F4A-C is not missing, or if filter variable indicates person raised livestock, missing data is imputed where available. The Household mean is used first if at least one other HH member reported livestock data for that variable. If the value was still missing, the Community or County mean values were used if at least 3 other individuals in the community or city/county reported livestock data for that variable. If the value is still missing, all livestock work is allocated equally to women aged 21-60 in the HH.

Non-Retirement Wages

Variable: INDWAGE--Total individual income from all non-retirement wages earned by individuals. INDWAGE is a component of INDINC, total individual net income.

Data Files: H07WAGES—Income from Wages (one record per job)
H07JOBS—Occupations (one record per person of working age)

Basic Algorithm: Annual wage is calculated for each job record in the wages file. Generally, annual wage income is Months Worked times Average Monthly non-Retirement Wage, annualized, plus Bonuses and Other Cash or In-Kind Income. For 1989, annualized income from piece work is calculated. (This is the same algorithm used for HH income from non-retirement wages, except that income is not summed across individuals in the household.)

Source: C3, months worked last year (job level), 1991 - 2006
C8, average month's wages (job level), 1991 - 2006
I19, value of bonuses received last year (job level), 1989-2006
I101, other cash income (job level), 2006
I103, value of other non-cash income (job level), 2006
B2, B3B, B4, B5, B9, B10, filter questions (person level)
(B2D and J5 are used to calculate INDRET separately from this program.)

Imputation

If the person appears to be working, i.e., they report a job code (1991, 1993, 1997, 2000), or if they report working (in 2004 - 2006), then Months Worked and Salary are imputed if necessary. Data are not imputed for 1989, since the data structure does not lend itself to imputation.

If Months Worked is missing, 12 months is assumed.

If Salary is missing, then Salary is imputed from adjacent waves where the job is the same. Specifically, if the job has not changed and both adjacent waves are available, the Salary from those waves is averaged. If only one adjacent wave is available and the job is unchanged, the Salary for the adjacent wave is used. Separate processing is done for primary and secondary jobs. If no data is available from adjacent waves (or if the job has changed) then community means are used to impute salary where at least 3 values are available to average. Otherwise the county means are used if three values are available. Although filter variables are available for I19, I101, and I103, these incomes were not imputed since the types of income are thought to be too irregular.

Retirement Wages

Variable: INDRET - Total Individual Retirement Income, one component of INDINC, Total Individual Income.

Data Files: M07OINC (1989 through 2004)
H07WAGES (2004 and 2006)
C07MAST (for interview date)

Note: In all Surveys through 2004, Retirement Income was reported under Other Income M07OINC (variable J5--Household Income from Retirement Pensions or Salaries in last 12 months.) In 2004, another question was also asked about Retirement Income and stored in M07WAGES (variable B2D--Avg. Monthly Retirement Wage from this job last year.) In 2006, J5 was dropped and B2D was retained.

Source: J5, retirement pensions/salaries (individual), 1989 - 2000
B2D, retirement wage from this job (job level), 2004 - 2006

Algorithm:

For 1989 - 2000, J5 is available for Household annual retirement income. If only one person retired, this income is assigned to that person. If more than one person retired, the income is divided proportionate to their wages in current or previous year, if available, or equally to women ages 55+ and men ages 60+ (government retirement age during those years).

For 2004 and following, the annual retirement income for each job is calculated from B2D times the # of months in the past year the person was retired from the job (based on retirement date and interview date). Note that where length of retirement cannot be calculated, 12 months are assumed. Calculated annual retirement income for each job is then aggregated to the individual level if an individual has retirement income from more than one job.

Imputation

None. Before 2004, no imputations can be done for J5 as there is no filter variable. Starting in 2004, B2D is available and is assumed to indicate retirement. (There were no cases in 2004 or 2006, where B2A indicated retirement but B2D was missing.) Therefore no code has been written to impute data. If problems with missing values for B2D arise in the future, code for imputing B2D may be added.

Total Individual Income, Nominal

Variable: INDINC - Total net individual income, nominal.
INDINCimp - Some element of INDINC is imputed.

Data files: INDBUSNyyyy - business income
INDFARMyyyy - farming income
INDFISHyyyy - fishing income
INDGARDyyyy - gardening income
INDLVSTyyyy - livestock income
INDRETIREyyyy - retirement income
INDWAGEyyyy - non-retirement wages

Source:	INDBUS	INDBUSimp
	INDFARM	INDFARMimp
	INDFISH	INDFISHimp
	INDGARD	INDGARDimp
	INDLVST	INDLVSTimp
	INDRET	INDRETimp
	INDWAGE	INDWAGEimp

Basic Algorithm: Sums income from all sources (above) and consolidates imputation flags for all waves of data. Concatenates all waves into one file. If all seven components of income are missing, indinc is missing. If any component of indinc is imputed, indincimp=1.

Total Individual Income, Inflated to 2006

Variable: INDINC_CPI

Data Files: ALLINDINC—total individual income, nominal, all waves
CPI_CHINA88_06_FINAL—inflation indexes, 1988 - 2006

Source: INDINC, total household income, nominal
INDEXURBAN_NEW, inflation index for urban areas, 1988 - 2006
INDEXRURAL_NEW, inflation index for rural areas, 1988 - 2006

Contents of File C07INDINC

HHID	MOST CURRENT HOUSEHOLD ID ON C05MAST
LINE	LINE NUMBER: IN A 2006 HOUSEHOLD
WAVE	Survey Year
COMMID	COMMUNITY ID: T1-T4
T1	PROVINCE
urban	1=Urban, 0=Rural
indbusimp	Some element of Indiv Business Work is Imputed
indbus	Individual Business Income
indfarmimp	Some element of Indiv farming Income is Imputed - 1989
indfarm	Individual Farming Income
indfishimp	Some element of Indiv Fishing Income is Imputed - 1989
indfish	Individual fishing Income
indgardimp	Some element of Indiv Gardening Income is Imputed - 1989
indgard	Individual Gardening Income
indlvstimp	Some element of Indiv livestock Income is Imputed - 1989
indlvst	Individual Livestock Income
indretimp	Some elements of retirement income imputed
indret	Individual retirement income
indwageimp	Some element(s) of indwage is imputed
indwage	Annual wage, bonus, other inc this job
indincimp	Some elements of indinc are imputed
indinc	Total nominal individual income
index_new	Inflation index to 2006
index_old	Deflation index to 1988
indinc_cpi	Total Individual income inflated to 2006

Basic Algorithm: divide household income by the constructed consumer price index (inflation index). The procedure used to construct the consumer price index (CPI) is documented in "Household Income Variable Construction.doc".