DSDR | DATA SHARING FOR DEMOGRAPHIC RESEARCH

ICPSR 3012

Chinese Household Income Project, 1995

Carl Riskin

Zhao Renwei

Li Shi

Original Primary Investigator Estimating Household Income

Terms of Use

The terms of use for this study can be found at:
http://www.icpsr.umich.edu/cocoon/ICPSR/TERMS/3012.xml

Information about Copyrighted Content

Some instruments administered as part of this study may contain in whole or substantially in part contents from copyrighted instruments. Reproductions of the instruments are provided as documentation for the analysis of the data associated with this collection. Restrictions on "fair use" apply to all copyrighted content. More information about the reproduction of copyrighted works by educators and librarians is available from the United States Copyright Office.

To: Li Shi/Zhao Renwei, Carl Riskin, Marc Eichen, Keith Griffin,
        John Knight and Mark Brenner

From: A.R. Khan

<u>Estimating Household Income</u>

This note provides the definitions of rural and urban household
income for the 1995 survey. It tries to conform to the categories
for 1988 although changes in China's economic organization have
eliminated some categories and the improvement in the survey
questionnaire has made it possible to achieve disaggregation of
certain other categories. Numbers refer to the serial numbers of
questions in the English translations made by Carl Riskin of rural
and urban questionnaires. For the rural questionnaire we require
some additional labelling: for items 501, 502, 901, 903 and 904
individual columns are labelled X, Y, Z and W in that order (note
that not all items have four columns). For questions with more than
one possible answer (e.g., question 806) the answers are labelled
X, Y, Z etc. Thus 901aX means "household consumption in 1995 (in
kg) of self-produced <u>wheat</u>"; 901aY means "household consumption in
1995 (in kg) of self-produced <u>rice</u>", etc.; 806X means "purchase
price of self-owned house"; and 806Y means "present (i.e., at 1995
price) value of self-owned house".

<u>General rule in case of a missing value</u>: In case of a missing value
of a variable the general rule is that, unless we can argue that
the value is genuinely zero, use the county, province or China
average in that order.

<u>Rural Household Income</u>

For the rural households the following components of income were
identified for the 1988 study:

      RY = RY1 + RY2 + RY3 + RY4 + RY5 + RY6 + RY7 + RY8

where

        RY = Disposable rural household income
        RY1 = Income from wages pensions and other compensations
              received by <u>individual members</u> of the household
        RY2 = Household income from township, village, collective and
              other types of enterprise (other than compensation for
              labor)

RY3 = Cash income from farming and industrial and subsidiary
            activities
      RY4 = Gross value of self-consumption of farm products
      RY5 = Income from property
      RY6 = Rental value of housing equity
      RY7 = Net transfer from/to collective and state entities
      RY8 = Miscellaneous income (including private transfer)

For 1995 we have all the above categories and the following
additional ones:

      RY3A = Net cash income from the sale of farm products
      RY3B = Net income from non-farm subsidiary activities
      RY3C = RY3A + RY4 = Net income from farming[1]

We show below the items in the questionnaire of which each of the
above sources of income is composed of.

RY1: Income from wages, pensions and other compensations

This component is the sum of such payments received by all the
individual members of the household. Thus we need to aggregate the
information for all individual members listed against each
question.

Operation	Variable	Explanation and comments
Add	201	Income of retired members (pensions, re-employment income & subsidies)
Add	12 times 202	Monthly average regular wage (including bonus, subsidies etc.) multiplied by 12
Add	203	Total non-regular income from work unit
Add	204	Cash income from other channels
Add	205	Income in kind
Add	207	Income from serving as village cadre
Add	208	Other cash income except from family operations
Add	209	Income brought back by members who have been temporarily working outside

---

[1] It should be noted that our definition of RY3C suffers
from the problem that production is not necessarily equal to the
sum of the quantity sold and the quantity self-consumed. The
difference between this sum and production is the change in stock
for which we have no information. We had the same problem in
1988.

Riskin thinks that 209 is included in 202 through 208. I find this interpretation a little difficult to accept, no doubt due to my inability to access the Chinese version of the questionnaire. Please advise.

RY2: Household income from township, village, collective and other types of enterprises other than compensation for labor

Add   206   Net individual income from private & other
                        forms of enterprise

Add   503   Income from collective welfare fund

RY3: Household cash income from the sale of the products of farming and industrial and subsidiary activities

Add   RY3A   See below

Add   RY3B   See below

RY3A: Household cash income from the sale of farm products

Add   501X   Gross income from household farming
                        operations

Subtract   501Y   Costs of production (including labor) of
                        farming operations

Subtract   717   Land contract fees

Note that RY3A can sometimes be negative without implying a negative return from farming operations. This is because 501Y and 717 are payments for the production of entire RY3C which consists of RY3A and RY4.

RY3B: Household net income from non-farm subsidiary activities

Add   502X   Gross income from industrial and
                        subsidiary activities

Subtract   502Y   Costs of production (including labor) of
                        non-farm subsidiary operations

RY4: Gross value of self-consumption of farm products

                                     Market value of the consumption
                                        of self-produced:

Add 901aX multiplied by 903aX   Wheat

Add 901aY multiplied by 903aY   Rice

```
Add  901aZ multiplied by 903aZ    Other grain (corn)

Add  904aX multiplied by 904aY    Meat

Add  904bX multiplied by 904bY    Seafood

Add  904cX multiplied by 904cY    Eggs

Add  904dX multiplied by 904dY    Edible oil

Add  904eX multiplied by 904eY    Sugar

Add  904fX multiplied by 904fY    Vegetables

Add  904gX multiplied by 904gY    Fruits and melons
```

We do not need to subtract current inputs from RY4 because these current inputs are included in the current inputs for RY3A.

<u>RY5: Household income from property</u>

```
Add        507          Income from renting/contracting out land

Add        508          Income from rental of other property

Add        509          Interest & dividend income
```

<u>RY6: Rental value of housing equity</u>

```
Add  0.08 times 806Y    Rental value of housing is assumed to be
                        8 percent of the current market value of
                        the house

Subtract 0.08 times 807a Interest on housing debt
```

<u>Note: RY6 must be non-negative, i.e., if the equity turns out to be negative we must ignore it and make the rental value of housing equity = 0.</u>

<u>RY7: Net transfer from state and collective</u>

```
Add        503b         Relief payments

Add        504c         Subsidies for the aged

Add        708b         Medical expenses paid by collective or
                        government. Although this item shows up
                        under household expenditure we have
```

		included it as it seems certain that it is not included anywhere else.
Subtract	714	Taxes on primary sector activities
Subtract	715	Taxes on secondary sector activities
Subtract	716	Taxes on tertiary sector activities
Subtract	718	Payment to collective

Note: 714 through 716 are treated as direct taxes. They may include indirect taxes which should properly have been subtracted from RY3A and RY3B. But this is not possible because we can not separate indirect taxes from direct taxes. Item 718 should clearly have been treated as cost of production. This again is not done because it can not be divided into separate payments for RY3A and RY3B. Our treatment of this is a departure from the 1988 definition (which did not require a separation of payment for RY3A from payment for RY3B). It does not appear that this makes a quantitatively significant difference.

RY8: Other income (including private transfer)

Add	504a	Fee for participating in survey
Add	505	Remittances by household members who are floating migrants
Add	506	Income/gift from friends and relatives
Add	510	Other income

Note: Annual net household income shown in 601 - 603

As was the case with the comparable item (HNET88) in the 1988 questionnaire, we believe that these are based on the State Statistical Bureau's definition of rural income.

Urban Household Income

For urban households income was divided into the following ten components for the 1988 study:

$$UY = UY1 + UY2 + UY3 + UY4 + UY5 + UY6 + UY7 + UY8 + UY9 + UY10$$

where

```
UY   = Disposable urban household income
UY1  = Cash income of the working members
UY2  = Income of the retired members
UY3  = Income of the non-working members
UY4  = Income from private/individual enterprises
UY5  = Income from property
UY6  = Miscellaneous income (including private transfer and
       special income)
UY7  = Subsidies less taxes (except housing subsidy and ration
       coupon subsidy) and income in kind
UY8  = Ration coupon subsidy
UY9  = Housing subsidy
UY10 = Rental value of owner occupied housing equity
```

For 1995 UY3 (which in 1988 mainly consisted of price subsidies) and UY8 do not exist any more due to the reform of public distribution and price subsidies. These categories will therefore be excluded for 1995. The notations for the rest will however be kept unchanged for easy comparability with 1988.

For UY1 through UY7 each category is the sum of such payments received by all the individual members of the household.

The contents of each component of UY in terms of the items enumerated in the questionnaire are as follows (the number shown in the column headed "variable" refers to the corresponding item in the urban questionnaire):

UY1: Cash income of the working members

Operation	Variable	Explanation and comments
Add	202	Wage including bonus, allowances & subsidies, overtime & special wages
Add	203	Other income from work unit (including hardship allowance)
Add	206	Wage income of employees of individual enterprises
Add	208	Other employee income
Add	209	Other labor income (including income from second job)

UY2: Income of the retired members

Add	207	Income of re-employed retired members
Add	211 (1)	Pensions

<u>UY4: Income from private/individual enterprises (non-wage)</u>

Add   204   Private enterprise proprietor's pre-tax
                          income

Add   205   Individual enterprise proprietor's pre-tax
                          income

Add   212   Income from household sideline production

<u>UY5: Income from property</u>

Add   210   Total property income

<u>UY6: Miscellaneous income (private transfer, special income etc.)</u>

Add   211 (3)   Alimony

Add   211 (4)   Money gifts

Add   211 (5)   Boarding fees paid by friends & relatives

Add   211 (6)   Payments for keeping accounts for the
                          survey

<u>UY7: Non-housing subsidies (including income in kind) less taxes</u>

Add   216   Total net income in kind

Add   211 (2)   Price subsidies

Add   211 (8)   Relief payments/hardship subsidies from
                          government

Subtract   213   Individual income taxes

<u>UY9: Housing subsidy: Enter only if answer to 608 is 1 or 2 (i.e.,
if housing is publicly-owned)</u>

Add   613 times 12   Monthly <u>market</u> rent multiplied by 12

Subtract   609 times 12   Actual monthly rent multiplied by 12

<u>If UY9 turns out to be negative, put it equal to zero</u>.

<u>UY10: Rental value of owner occupied housing: Enter only if answer
to 608 is 3,4 or 5 (i.e., if the house is owned by the occupant)</u>

Add   613 times 12   Monthly <u>market</u> rent multiplied by 12

Subtract   0.08 times 406 (1)   Eight per cent interest on housing
                                 debt

Note that UY10 must be non-negative, i.e., if interest on housing debt exceeds annual market rent, we must put UY10=0.

Note: Item 201 is income according to State Statistical Board definition, comparable to 601-603 in the rural questionnaire.

(End of Note)

Estimating Rural Household Income
SAS Code

/* This is a program designed to calculate the components of
rural household income for mainland China, using the datasets
Ruralh and Ruralp */

* ==================== part one ==================== *;

/* Part one: (Module one) Set initial dataset */

/* Part one: (Module two) Create values for the variables that are
missing that cannot be allowed to be equal to zero. These variables
are prices for self-consumed agricultural products (b903a_1 b903a_2
b903a_3 b904a_2 b904b_2 b904c_2 b904d_2 b904e_2 b904f_2 b904g_2)
rental value of housing (b806_2), relief payments (b504b),
and the size of house in square metres (b1002) */

/* Part one: (Module three) Set the erroneous prices
equal to county average */

* ==================== module one ==================== * ;

/* USERS NEED TO MODIFY THE LIBNAME STATEMENT TO SUIT THEIR
CONFIGURATIONS */

libname china95 'd:\sasdata\china95';

/* create temporary dataset=aa in which are located all the variables
of interest as well as province, county and household id */

data aa; set china95.ruralh(keep=a1 b101 province county b903a_1
nhh b903a_2 b903a_3 b904a_2 b904b_2 b904c_2 b904d_2 b904e_2 b904f_2
b904g_2 b806_2 b504b b1002 b901_1 b901_2 b901_3 b901a_1 b901a_2
b901a_3);
dummy=1;

* ==================== end module I ==================== * ;
* ==================== module II ==================== * ;

/* begin by setting all the zeros equal to missing */

Array AA (I) b903a_1 b903a_2 b903a_3 b904a_2 b904b_2
b904c_2 b904d_2 b904e_2 b904f_2 b904g_2 b806_2 b504b b1002;
do over AA ; if AA=0 then AA=.; end;

/* then calculate the county average */

```
proc means data=aa mean noprint; var b903a_1 b903a_2 b903a_3 b904a_2
b904b_2 b904c_2 b904d_2 b904e_2 b904f_2 b904g_2 b806_2 b504b b1002;
output out=a1 mean=b903a_1m b903a_2m b903a_3m b904a_2m
b904b_2m b904c_2m b904d_2m b904e_2m b904f_2m b904g_2m b806_2m
b504bm b1002m;
by province county;

/* then calculate the province average */

proc means data=aa mean noprint; var b903a_1 b903a_2 b903a_3 b904a_2
b904b_2 b904c_2 b904d_2 b904e_2 b904f_2 b904g_2 b806_2 b504b b1002;
output out=a2 mean=pb903a_1 pb903a_2 pb903a_3 pb904a_2
pb904b_2 pb904c_2 pb904d_2 pb904e_2 pb904f_2 pb904g_2 pb806_2 pb504b
pb1002;
by province;

/* then calculate the country average */

proc means data=aa mean noprint; var b903a_1 b903a_2 b903a_3 b904a_2
b904b_2 b904c_2 b904d_2 b904e_2 b904f_2 b904g_2 b806_2 b504b b1002;
output out=a mean=db903a_1 db903a_2 db903a_3 db904a_2
db904b_2 db904c_2 db904d_2 db904e_2 db904f_2 db904g_2 db806_2 db504b
db1002;
by dummy;

/* merge them all together */

data a3; set aa; run;
data b; merge a3 a1; by province county; run;
data b; merge b a2; by province; run;
data b2; merge b a; by dummy; run;
data c; set b2;

/* replace the values by the various means */

if b903a_1=. and b903a_1m>0 then b903a_1=b903a_1m;
else if b903a_1=. and pb903a_1>0 then b903a_1=pb903a_1;
else if b903a_1=. then b903a_1=db903a_1;

if b903a_2=. and b903a_2m>0 then b903a_2=b903a_2m;
else if b903a_2=. and pb903a_2>0 then b903a_2=pb903a_2;
else if b903a_2=. then b903a_2=db903a_2;

if b903a_3=. and b903a_3m>0 then b903a_3=b903a_3m;
else if b903a_3=. and pb903a_3>0 then b903a_3=pb903a_3;
```

```
else if b903a_3=. then b903a_3=db903a_3;

if b904a_2=. and b904a_2m>0 then b904a_2=b904a_2m;
else if b904a_2=. and pb904a_2>0 then b904a_2=pb904a_2;
else if b904a_2=. then b904a_2=db904a_2;

if b904b_2=. and b904b_2m>0 then b904b_2=b904b_2m;
else if b904b_2=. and pb904b_2>0 then b904b_2=pb904b_2;
else if b904b_2=. then b904b_2=db904b_2;

if b904c_2=. and b904c_2m>0 then b904c_2=b904c_2m;
else if b904c_2=. and pb904c_2>0 then b904c_2=pb904c_2;
else if b904c_2=. then b904c_2=db904c_2;

if b904d_2=. and b904d_2m>0 then b904d_2=b904d_2m;
else if b904d_2=. and pb904d_2>0 then b904d_2=pb904d_2;
else if b904d_2=. then b904d_2=db904d_2;

if b904e_2=. and b904e_2m>0 then b904e_2=b904e_2m;
else if b904e_2=. and pb904e_2>0 then b904e_2=pb904e_2;
else if b904e_2=. then b904e_2=db904e_2;

if b904f_2=. and b904f_2m>0 then b904f_2=b904f_2m;
else if b904f_2=. and pb904f_2>0 then b904f_2=pb904f_2;
else if b904f_2=. then b904f_2=db904f_2;

if b904g_2=. and b904g_2m>0 then b904g_2=b904g_2m;
else if b904g_2=. and pb904g_2>0 then b904g_2=pb904g_2;
else if b904g_2=. then b904g_2=db904g_2;


if b806_2=. and b806_2m>0 then b806_2=b806_2m;
else if b806_2=. and pb806_2>0 then b806_2=pb806_2;
else if b806_2=. then b806_2=db806_2;

if b504b=. and b504bm>0 then b504b=b504bm;
else if b504b=. and pb504b>0 then b504b=pb504b;
else if b504b=. then b504b=db504b;

if b1002=. and b1002m>0 then b1002=b1002m;
else if b1002=. and pb1002>0 then b1002=pb1002;
else if b1002=. then b1002=db1002;

* ====================end of module II ==================;
/* module III supplement */
```

if b903a_1m=. then b903a_1m=pb903a_1; else b903a_1m=b903a_1m;
if b903a_2m=. then b903a_2m=pb903a_2; else b903a_2m=b903a_2m;
if b903a_3m=. then b903a_3m=pb903a_3; else b903a_3m=b903a_3m;
if b904a_2m=. then b904a_2m=pb904a_2; else b904a_2m=b904a_2m;
if b904b_2m=. then b904b_2m=pb904b_2; else b904b_2m=b904b_2m;
if b904c_2m=. then b904c_2m=pb904c_2; else b904c_2m=b904c_2m;
if b904d_2m=. then b904d_2m=pb904d_2; else b904d_2m=b904d_2m;
if b904e_2m=. then b904e_2m=pb904e_2; else b904e_2m=b904e_2m;
if b904f_2m=. then b904f_2m=pb904f_2; else b904f_2m=b904f_2m;
if b904g_2m=. then b904g_2m=pb904g_2; else b904g_2m=b904g_2m;

if b903a_1m=. then b903a_1m=db903a_1; else b903a_1m=b903a_1m;
if b903a_2m=. then b903a_2m=db903a_2; else b903a_2m=b903a_2m;
if b903a_3m=. then b903a_3m=db903a_3; else b903a_3m=b903a_3m;
if b904a_2m=. then b904a_2m=db904a_2; else b904a_2m=b904a_2m;
if b904b_2m=. then b904b_2m=db904b_2; else b904b_2m=b904b_2m;
if b904c_2m=. then b904c_2m=db904c_2; else b904c_2m=b904c_2m;
if b904d_2m=. then b904d_2m=db904d_2; else b904d_2m=b904d_2m;
if b904e_2m=. then b904e_2m=db904e_2; else b904e_2m=b904e_2m;
if b904f_2m=. then b904f_2m=db904f_2; else b904f_2m=b904f_2m;
if b904g_2m=. then b904g_2m=db904g_2; else b904g_2m=b904g_2m;

* ====================module III ===================== ;
if (.5*b903a_1>b903a_1m or b903a_1m>2*b903a_1) then b903a_1=b903a_1m;
else b903a_1=b903a_1;

if (.5*b903a_2>b903a_2m or b903a_2m>2*b903a_2) then b903a_2=b903a_2m;
else b903a_2=b903a_2;

if (.5*b903a_3>b903a_3m or b903a_3m>2*b903a_3) then b903a_3=b903a_3m;
else b903a_3=b903a_3;

if (.5*b904a_2>b904a_2m or b904a_2m>2*b904a_2) then b904a_2=b904a_2m;
else b904a_2=b904a_2;

if (.5*b904b_2>b904b_2m or b904b_2m>2*b904b_2) then b904b_2=b904b_2m;
else b904b_2=b904b_2;

if (.5*b904c_2>b904c_2m or b904c_2m>2*b904c_2) then b904c_2=b904c_2m;
else b904c_2=b904c_2;

if (.5*b904d_2>b904d_2m or b904d_2m>2*b904d_2) then b904d_2=b904d_2m;
else b904d_2=b904d_2;

if (.5*b904e_2>b904e_2m or b904e_2m>2*b904e_2) then b904e_2=b904e_2m;
else b904e_2=b904e_2;

```
if (.5*b904f_2>b904f_2m or b904f_2m>2*b904f_2) then b904f_2=b904f_2m;
else b904f_2=b904f_2;

if (.5*b904g_2>b904g_2m or b904g_2m>2*b904g_2) then b904g_2=b904g_2m;
else b904g_2=b904g_2;




/* get rid of all the averages */

drop b903a_1m b903a_2m b903a_3m b904a_2m
b904b_2m b904c_2m b904d_2m b904e_2m b904f_2m b904g_2m b806_2m
b504bm b1002m pb903a_1 pb903a_2 pb903a_3 pb904a_2
pb904b_2 pb904c_2 pb904d_2 pb904e_2 pb904f_2 pb904g_2 pb806_2 pb504b
pb1002 db903a_1 db903a_2 db903a_3 db904a_2
db904b_2 db904c_2 db904d_2 db904e_2 db904f_2 db904g_2 db806_2 db504b
db1002 dummy I;
run;

/* part one is finished and we have a dataset=c with the
following varibles (a1 a2 b101 b903a_1 b903a_2 b903a_3 b904a_2
b904b_2 b904c_2 b904d_2 b904e_2 b904f_2 b904g_2 b806_2 b504b
b1002 RN b901_1 b901_2 b901_3 b901a_1 b901a_2 b901a_3) */

* ===================== end module III =================== *;

* ===================== end of part one ================== * ;

* ===================== part two ======================= *;

/* part two will calculate the household sums of all the
variables which are individual specific, and will generate
a dataset with those household sums in it */

data e; set china95.ruralp(keep=a1 b101 b201 b202 b203 b204
b204a b204b b205 b206 b207 b208);
yb202=12*b202;
if (b204a+b204b)>b204 then b204=(b204a+b204b);
else b204=b204;

run;
proc means data=e sum noprint; var b201 yb202 b203 b204 b205 b206
b207 b208;
output out=f sum=sb201 syb202 sb203 sb204 sb205 sb206 sb207 sb208;
```

by a1 b101; run;

/* part two is finished and we have a dataset=f which has the
following variables (sb201 syb202 sb203 sb204 sb205 sb206 sb207
sb208) */

* ================= end of part two ==================== *;

* ===================== part three ====================== *;

/* in this section I will do two things. First, I will
extract the remaining household level data that I need to
calculate income. Then I will combine that with the two
previously constructed datasets to get a new dataset with
all the individual components of household income */

data g; merge china95.ruralh(keep=a1 b101 b412 b503 b501_1
b501_2 b717 b502_1 b502_2 b806_4 b904a_1 b904b_1 b904c_1 b904d_1
b904e_1 b904f_1 b904g_1 b507 b508 b509 b807a b504c b708b b714
b715 b716 b718 b504a b505 b506 b510)
c f; by a1 b101;
run;
/* now that is done and we have an output=g which has all
the varibles needed to calculate household income */

* ================= end part three ==================== *;

* ===================== part four ====================== *;

/* in this section I will manipulate the variables to calculate
RY1, RY2, RY3A, etc. as well as create a variable for the value
of self-consumed items. */

/* first I will calculate the value of self consumed products */

data kk; set g;

vb901a_1=b901a_1*b903a_1;
vb901a_2=b901a_2*b903a_2;
vb901a_3=b901a_3*b903a_3;
vb904a_1=b904a_1*b904a_2;
vb904b_1=b904b_1*b904b_2;
vb904c_1=b904c_1*b904c_2;
vb904d_1=b904d_1*b904d_2;
vb904e_1=b904e_1*b904e_2;
vb904f_1=b904f_1*b904f_2;

```
vb904g_1=b904g_1*b904g_2;

/* then I will calculate the rental value of home equity and
the interest on housing debt */

rb806_4=.08*b806_4;
rb807a=.08*b807a;

run;

data kk; set kk;

/* now I will calculate the various income components */


RY1=sb201+syb202+sb203+sb204+sb205+sb207+sb208;

RY2=sb206+b503;

RY3A=b501_1-b501_2-b717;

RY3B=b502_1-b502_2;

RY4=vb901a_1+vb901a_2+vb901a_3+vb904a_1+vb904b_1+
vb904c_1+vb904d_1+vb904e_1+vb904f_1+vb904g_1;

RY3=RY3A+RY3B-RY4;

RY5=b507+b508+b509;

RY6=rb806_4-rb807a;
if RY6<0 then RY6=0;

RY7=b504b+b504c+b708b-b714-b715-b716-b718;

RY8=b504a+b505+b506+b510;

RY=RY1+RY2+RY3A+RY3B+RY5+RY6+RY7+RY8;

RYPC=RY/nhh;



drop b901a_1 b901a_2 b901a_3 b904a_1 b904b_1 b904c_1 b904d_1
b904e_1 b904f_1 b904g_1
b903a_1 b903a_2 b903a_3 b904a_2 b904b_2 b904c_2 b904d_2 b904e_2
```

```
b904f_2 b904g_2
b806_2 b807a
_type_ _freq_;
run;

/* USERS NEED TO UNCOMMENT THE CODE BELOW TO GENERATE A PERMANENT
SAS DATASET FOR
RURAL INCOME */

/* data china95.rincome; set kk;
run;*/
```

Estimating Urban Household Income
SAS Code

```
/* USERS WILL NEED TO MODIFY THIS LIBNAME STATEMENT TO FIT LOCAL
OPERATING ENVIRONMENT */

libname china95 'd:\sasdata\china95';

/* Part One: Create household totals for income definition from individual values */

data a; set china95.urbanp(keep=n1 a52 a53 a54 a55 a56 a57 a58 a59
a60 a61 a62 a66 a68 a69 a67 province county
a77 a64 a65 a85 a71 a72 a73 a74 a75
a79 a80 a81 a82 a89 a90 a91 a92 a93 a94 a78 a84 a86);

Array AA (I) a52 a53 a54 a55 a56 a57 a58 a59
a60 a61 a62 a66 a68 a69 a67 a77 a79 a80 a81 a82 a89
a90 a91 a92 a93 a94 a78 a84 a86;
do over AA; if AA<0 then AA=0; end;

Array BB (I) a64 a65 a85 a71 a72 a73 a74 a75 ;
do over BB; if BB=. then BB=0; end;

if (a56+a57+a58+a59)>a55 then a55=(a56+a57+a58+a59);
if (a53+a54+a55+a60+a61)>a52 then a52=(a53+a54+a55+a60+a61);
if (a72+a73+a74+a75)>a71 then a71=(a72+a73+a74+a75);
if (a90+a91+a92+a93+a94)>a89 then a89=(a90+a91+a92+a93+a94);

proc means data=a sum noprint; var a52 a62 a66 a68 a69 a67
a77 a64 a65 a85 a71 a79 a80 a81 a82 a89 a78 a84 a86;
output out=a2 sum=a52h a62h a66h a68h a69h a67h
a77h a64h a65h a85h a71h a79h a80h a81h a82h a89h a78h a84h a86h;
by n1; run;

/* Part Two: Combine the individual totals with the remaining household variables
needed for
income definition */

data e; merge china95.urbanh(keep=province county nhh n1 h63 h68
h64 h14) a2;
by n1; run;
data kk; set e;
UY1=a52h+a62h+a66h+a68h+a69h;
UY2=a67h+a77h;
UY4=a64h+a65h+a85h;
UY5=a71h;
UY6=a79h+a80h+a81h+a82h;
UY7=a89h+a78h+a84h+a86h;
if(h63=1 or h63=2) then UY9=12*h68-12*h64;
```

```
else UY9=0;
if UY9<0 then UY9=0;
if (h63=3 or h63=4 or h63=5) then UY10=12*h68-.08*h14;
else UY10=0;
if UY10<0 then UY10=0;
UY=UY1+UY2+UY4+UY5+UY6+UY7+UY9+UY10;
UYPC=UY/NHH;
drop _type_ _freq_ ;
run;

/* USERS WILL NEED TO REMOVE THE COMMENTS BELOW IN ORDER TO SET A
PERMANENT SAS DATASET */

/* data china95.uincome; set kk;
run; */
```