

Determination of household wealth index

Our study participants were from both urban and rural areas. Therefore, the determination of their household wealth index should take this difference into account as the variables (assets) used to measure wealth may vary for urban and rural dwellers. In these regard, out of the 11 items we considered in measuring wealth index, one item (owning cattle) seem more common for rural and source of energy (especially electricity) seems more common for urban residents. This is just simply from our judgment, not from the statistics. All the remaining variables seem common in both settings. Therefore, we have conducted the following steps to determine the household wealth index using Stata.

Step 1: First, change all the 11 variables into binary as yes (present) or no (absent) that were coded as '1' and '0', respectively. This is because wealth index works better in binary variables. The variables type should also be 'scale' but not 'string' format.

Step 2: Check the frequency of each item (variable) both separately for urban and rural, and by merging them together. If the frequency is between 5% and 95%, it is acceptable to be included in the analysis. Otherwise, if it is rare (less than 5%) or more common (greater than 95%), then such variables are not helpful to differentiate persons using wealth index and thus should be excluded.

Therefore, first we check the frequency (percent) of each item for urban dwellers. Fortunately, all were more than 5% including cattle owning. The same result obtained for rural dwellers. Even the frequency is between 5% and 95% for both urban and rural migrants mixed together.

For example, the Stata command to determine frequency (percent) for those who own radio (if the variable name for radio is 'radio' in Stata) is:

```
tab radio
```

Table 1: frequency of radio owning

Does the household currently own any radio?	Freq.	Percent	Cum.
0	216	32.19	32.19
1	455	67.81	100.00
Total	671	100.00	

The result in Table 1 show that 67.81% owned radio which is in between 5% and 95%. Therefore, it would be included in to the analysis. We did same for other variables.

Step 3: Principal Component Analysis (PCA). Because assets in urban could be different from that of rural, may be by worthiness or price, we have to consider weighting the items. There are different types of weighting. One is to consider their price. The other, which is commonly used and is simple, is to determine the wealth index for urban and rural separately and then merging (mixing) them together and report for both urban and rural.

We used the second approach for our study participants using the following procedures. First, suppose that the 11 items/indicators to measure wealth index are denoted in the Stata software as: I1, I2, I3, ..., I11 (Table 2).

Table 2: The 11 items used to measure wealth index

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	I11
1	1	0	1	1	1	0	0	0	0	1	0
2	1	1	0	0	0	0	0	0	0	0	0
3	0	1	1	1	1	0	1	1	0	1	0
4	1	1	1	0	1	0	1	1	1	1	0
5	1	1	1	1	1	0	1	0	0	1	0
6	1	1	1	0	0	0	0	1	1	1	0
7	1	1	1	0	1	0	0	1	1	1	1
8	1	1	1	1	1	0	1	1	1	1	1
9	0	1	0	0	1	0	1	0	0	1	0
10	1	1	0	0	1	0	0	1	0	1	0
11	1	1	1	0	1	0	1	0	0	1	0
12	0	1	1	1	1	1	0	1	1	0	1
13	1	1	1	0	1	0	1	1	1	1	0

Step 3.1: Sort participants by residence.

Suppose that the variable residence (urban/rural) is denoted as “residence” which was coded as rural = “1” and urban = “2”. Our sample size was 671, and 410 of them were rural residents and the remaining were urban residents.

To sort participants by residence (for example, listing rural residents first and urban residents next), use the command “sort”. This is simply to make the next steps easier. The Stata command is:

```
sort residence
```

This command arranges our data set in such a way that participants listed from 1 to 410 would be rural residents (code 1) and from 411 to 671 would be urban residents (code 2) (Table 3).

Table 3: Code for residence against role number

	age_traffi~g	sex	religion	Marital_st~s	Ethnicity	educ	residence	Former_addr~s	Your_zone
407	20	2	2	1	2	1	1	Oromia	Arissie
408	18	2	3	2	3	4	1	Debube	Hosahena
409	19	2	2	2	2	3	1	Amhara	Weldia
410	22	2	2	2	3	4	1	Debube	Arabie
411	30	2	2	1	2	3	2	Oromyia	Harar
412	19	2	2	2	3	2	2	SNNP	Hadyia
413	25	2	3	3	3	2	2	SNNP	Hosaena
414	17	2	2	1	1	4	2	Amhara	Oromia Z
415	22	2	1	2	3	2	2	Oromiya	Jimma
416	20	2	2	2	2	4	2	Oromo	Jimma
417	24	2	2	2	2	4	2	Oromo	Merabe W
418	27	2	1	1	2	2	2	Oromia	Jimma
419	49	?	?	?	1	3	2	Iddie Bk	Iddie Bk

Step 3.2: First, we should determine PCA for urban and rural residents separately. The command in Stata is as follows:

```
bysort residence: pca I1 I2 I3 I4 I5 I6 I7 I8 I9 I10 I11
```

Here, “bysort” helps us to get separate PCA for urban and rural; “residence” is the grouping variable in determining the wealth index separately for urban and rural residents;

“pca” orders Stata to conduct Principal Component Analysis. Therefore, 11 principal components will be generated for both urban and rural residents.

Step 3.2: To predict (produce) rural wealth index scores (let us denote it as rural_wis), use the following command:

```
predict rural_wis if residence==1
```

This command results that the variable “rural_wis” is created in Stata’s Data Editor Window spreadsheet just at the end of all variables or in the last column, and the scores would be generated for the first 410 participants (see step 3.1 above that these are rural residents) (Table 4).

Table 4: Newly generated wealth index scores for rural residents named as rural_wis

	I5	I6	I7	I8	I9	I10	I11	rural_wis
400	0	0	1	1	0	1	1	-.0304251
401	0	0	0	0	0	1	0	-3.180683
402	0	0	0	1	1	1	0	-2.332154
403	1	0	1	0	0	1	0	-1.137494
404	0	0	1	0	1	1	1	-1.122466
405	0	0	0	0	0	1	0	-3.803552
406	1	0	1	1	1	1	0	.4794715
407	0	0	0	0	0	1	0	-3.803552
408	0	0	0	1	1	1	0	-1.729188
409	1	0	1	1	1	1	0	-.3093706
410	1	0	1	0	0	1	0	-1.760363
411	1	1	1	1	1	0	1	.
412	0	0	0	0	0	1	1	.
413	1	0	1	1	0	1	0	.
414	1	0	1	0	1	0	0	.
415	1	0	1	1	1	0	1	.
416	1	1	1	1	1	0	1	.
417	1	0	1	0	1	1	1	.
418	1	1	1	1	0	1	0	.
419	1	1	1	1	1	0	1	.

Similarly, to generate for urban residents’ wealth index scores (let us denote it by urban_wis), use the following command:

```
predict urban_wis if residence==2
```

This command results that the variable “urban_wis” is created in another column next to rural_wis in Stata’s Data Editor Window spreadsheet. The scores would be generated from number 411 to 671 (see step 3.1 above that these are urban residents) (Table 5).

Table 5: Newly generated wealth index scores for urban residents named as urban_wis

	I5	I6	I7	I8	I9	I10	I11	rural_wis	urban_wis	w
400	0	0	1	1	0	1	1	-.0304251	.	
401	0	0	0	0	0	1	0	-3.180683	.	
402	0	0	0	1	1	1	0	-2.332154	.	
403	1	0	1	0	0	1	0	-1.137494	.	
404	0	0	1	0	1	1	1	-1.122466	.	
405	0	0	0	0	0	1	0	-3.803552	.	
406	1	0	1	1	1	1	0	.4794715	.	
407	0	0	0	0	0	1	0	-3.803552	.	
408	0	0	0	1	1	1	0	-1.729188	.	
409	1	0	1	1	1	1	0	-.3093706	.	
410	1	0	1	0	0	1	0	-1.760363	.	
411	1	1	1	1	1	0	1	.	2.741562	
412	0	0	0	0	0	1	1	.	-2.340302	
413	1	0	1	1	0	1	0	.	-.260398	
414	1	0	1	0	1	0	0	.	-.0788388	
415	1	0	1	1	1	0	1	.	1.638136	
416	1	1	1	1	1	0	1	.	2.741562	
417	1	0	1	0	1	1	1	.	.4422542	
418	1	1	1	1	0	1	0	.	.6770558	
419	1	1	1	1	1	0	1	.	2.741562	

Now, we have already generated rural and urban residents' wealth index scores separately. Note that these two variables (urban and rural wealth index scores) were generated separately (by their own scale or setting). Thus, we are not expected to weight them further since they were generated within each setting or scale. Therefore, we must create another variable for both urban and rural (a composite wealth index score) by merging the two newly generated scores.

Step 3.3: Let the composite wealth index score for both urban and rural be denoted as "wealth_index_score". Note that you can give any other name that you want provided that it is acceptable by Stata. We can create this variable manually easily. First open the Data Editor Window or spreadsheet. Just copy the rural wealth index scores generated under the variable "rural_wis" (i.e. from 1 -410) and paste under the wealth_index_score from 1-410. Similarly copy the scores generated under urban_wis from 411-671 and past under wealth_index_score from 411-671. Now, we have the new variable (a composite wealth index score) containing scores for both urban and rural (Table 6).

Table 6: Newly generated composite wealth index scores (for urban and rural) named as wealth_index_scores.

	I5	I6	I7	I8	I9	I10	I11	rural_wis	urban_wis	wealth_ind-e
400	0	0	1	1	0	1	1	-.0304251	.	-.0304251
401	0	0	0	0	0	1	0	-3.180683	.	-3.180683
402	0	0	0	1	1	1	0	-2.332154	.	-2.332154
403	1	0	1	0	0	1	0	-1.137494	.	-1.137494
404	0	0	1	0	1	1	1	-1.122466	.	-1.122466
405	0	0	0	0	0	1	0	-3.803552	.	-3.803552
406	1	0	1	1	1	1	0	.4794715	.	.4794715
407	0	0	0	0	0	1	0	-3.803552	.	-3.803552
408	0	0	0	1	1	1	0	-1.729188	.	-1.729188
409	1	0	1	1	1	1	0	-.3093706	.	-.3093706
410	1	0	1	0	0	1	0	-1.760363	.	-1.760363
411	1	1	1	1	1	0	1	.	2.741562	2.741562
412	0	0	0	0	0	1	1	.	-2.340302	-2.340302
413	1	0	1	1	0	1	0	.	-.260398	-.260398
414	1	0	1	0	1	0	0	.	-.0788388	-.0788388
415	1	0	1	1	1	0	1	.	1.638136	1.638136
416	1	1	1	1	1	0	1	.	2.741562	2.741562
417	1	0	1	0	1	1	1	.	.4422542	.4422542

Step 4: Produce wealth index. This can be done easily by grouping the composite wealth index scores into five equal parts and sometimes into three equal parts.

```
xtile wealth_index= wealth_index_score, nq(5)
```

Now, we will have a variable called wealth_index with values 1, 2, 3, 4, and 5. The higher the index, the wealthier the person is (Table 7).

Table 7: Newly generated composite wealth index (for rural and urban residents) named as wealth_index.

	I5	I6	I7	I8	I9	I10	I11	rural_wis	urban_wis	wealth_ind-e	wealth_index
400	0	0	1	1	0	1	1	-.0304251	.	-.0304251	4
401	0	0	0	0	0	1	0	-3.180683	.	-3.180683	1
402	0	0	0	1	1	1	0	-2.332154	.	-2.332154	1
403	1	0	1	0	0	1	0	-1.137494	.	-1.137494	2
404	0	0	1	0	1	1	1	-1.122466	.	-1.122466	3
405	0	0	0	0	0	1	0	-3.803552	.	-3.803552	1
406	1	0	1	1	1	1	0	.4794715	.	.4794715	5
407	0	0	0	0	0	1	0	-3.803552	.	-3.803552	1
408	0	0	0	1	1	1	0	-1.729188	.	-1.729188	2
409	1	0	1	1	1	1	0	-.3093706	.	-.3093706	3
410	1	0	1	0	0	1	0	-1.760363	.	-1.760363	2
411	1	1	1	1	1	0	1	.	2.741562	2.741562	5
412	0	0	0	0	0	1	1	.	-2.340302	-2.340302	1
413	1	0	1	1	0	1	0	.	-.260398	-.260398	4
414	1	0	1	0	1	0	0	.	-.0788388	-.0788388	4
415	1	0	1	1	1	0	1	.	1.638136	1.638136	5
416	1	1	1	1	1	0	1	.	2.741562	2.741562	5
417	1	0	1	0	1	1	1	.	.4422542	.4422542	5
418	1	1	1	1	0	1	0	.	.6770558	.6770558	5

If you want to divide it into three equal parts, use the following command:

```
xtile wealth_index= wealth_index_score, nq(3)
```

In this case, we will have a variable called wealth_index with values 1, 2, and 3.