Technical Note No. 5

Note on the Wealth Index

NOTE ON THE WEALTH INDEX

Background

The National Social Protection Monitoring Survey (NSPMS) wealth index has followed the guidelines provided in Issue 6 of the 2004 *DHS Comparative Reports* series (ORC Macro, 2004). This issue is particularly informative regarding the necessary steps to build a wealth index and also provides a cross-continental comparative perspective of its application.

Household wealth is seen as a measure of the economic status of the family in the absence of detailed and reliable information on income and expenditure. Its purpose is to rank households based on an underlying unobserved variable (wealth) that is assumed to be correlated with a set of indicator variables. These variables usually consist of assets owned by the households and services that they have access to. Information about these indicators is collected by most household surveys, especially the Demographic and Health Survey (DHS) and the Multiple Indicator Cluster Survey (MICS), which offer comparable data worldwide.

ORC Macro (2004) recommends that all household assets and utility services should be included in the calculation of the wealth index, since the larger the "number of indicator variables, the better the distribution of households with fewer households being concentrated on certain index scores" (ORC Macro, 2004: 8). In addition to using as much information as possible, it recommends that the different categories of an indicator variable should be used fully, without merging different categories into a single one, so that each category becomes a dichotomous variable (ORC Macro, 2004). The reason for this is that sometimes the ordering of the different

categories is not obvious, as it may depend on the context and location of the household.

The report also recommends the use of Principal Component Analysis as put forward by Filmer and Prichett (2001) to assign weights to the different indicator variables and their categories. The weights are essentially based on the correlation between the indicator variables and the unobserved wealth status as given by the factor coefficient scores (or factor loadings). Only the first factor produced is used to serve as proxy for the wealth index. For each household, the factor loadings are multiplied by the indicator values and summed up in order to yield the household's wealth index value.

The NSPMS Final report wealth index model

The wealth index quintile used in the NSPMS Final Report is calculated at the household level using variables observed in round 1. In a few cases, in order to deal with missing information for non-missing households in round 1, we had to impute the values reported in round 2 into the missing information for round 1. The wealth index was calculated using the "balanced weight" (adj_weight_w1ew2ew3ew4) since all indicators in the Final Report are calculated using the balanced sample.

The specific indicators used in the construction of the NSPMS wealth index are: floor materials; water access/source; types of toilet facilities (sanitation); access to electricity; house ownership; land ownership/possession; crowding; radio; TV; TV satellite receiver; mobile phone; telephone landline; bed; refrigerator; fan; bicycle; bus; car; motorcycle; truck; animal cart; sewing machine; water pump and tractor. They make up a broader set of indicators than the ones used in the MICS (2006),

which did not include information on floor material, water access, toilet facilities or crowding.

Table 1 shows the loading factors which gauge the correlation between the unobserved wealth and the indicator and varies according to the following interval (-1,1). Assets such as refrigerator, TV / TV satellite receiver and services such as no access to toilet facility, mud floor and toilet discharged into the sewage network are the variables with higher correlations — either positive or negative — with unobserved wealth as per their loading factors.

Table 1 Principal Component Analysis: Loading Factors

Variable	Loading factor	Uniqueness
Floor: concrete	0.2494	0.9378
Floor: tiles	0.5327	0.7162
Floor: mud/soil	-0.6445	0.5846
Floor: stone	0.0666	0.9956
Floor: marble	0.0387	0.9985
Floor: wood	0.0077	0.9999
Floor: other	0.0348	0.9988
Water: piped inside the dwelling	0.3981	0.8415
Water: piped inside the compound	-0.0983	0.9903
Water: public tap or standpipe outside the compound	-0.0937	0.9912
Water: tube well or borehole connected with pipes	-0.2474	0.9388
Water: protected dug well	-0.1253	0.9843
Water: unprotected dug well	-0.3169	0.8996
Water: protected spring	-0.0527	0.9972
Water: unprotected spring	-0.1547	0.9761

Water: rainwater harvesting/cistern	-0.0308	0.9991
Water: cart with small tank/drum	0.0099	0.9999
Water: tanker-truck	0.2091	0.9563
Water: surface water	-0.1926	0.9629
Water: bottled water	0.081	0.9934
Water: Jerry can filtered water	0.2768	0.9234
Water: other	-0.0286	0.9992
Toilet: flush or pour discharge into the sewage system	0.5355	0.7132
Toilet: flush or pour discharge into septic tank	-0.0319	0.999
Toilet: flush or pour latrine discharge into cesspit	0.2864	0.918
Toilet: ventilated improved pit	0.0122	0.9999
Toilet: pit latrine with slab as hole cover	-0.0185	0.9997
Toilet: pit latrine without slab as hole cover (open)	-0.0811	0.9934
Toilet: bucket	-0.0769	0.9941
Toilet: hanging toilet or latrine	-0.1063	0.9887
Toilet: no facility (bush/field)	-0.6388	0.592
Toilet: other	-0.011	0.9999
Toilet: flush latrine discharging to open	-0.0168	0.9997
Crowding	-0.3636	0.8678
Electricity	0.6718	0.5486
House ownership	-0.0226	0.9995
Land ownership/possession	-0.0335	0.9989
Radio	0.145	0.979
TV	0.7761	0.3977
TV satellite receiver	0.7614	0.4202
Phone: mobile	0.3414	0.8834
Phone: landline	0.4916	0.7583

Bed	0.4367	0.8093
Refrigerator	0.7438	0.4467
Fan	0.515	0.7348
Bicycle	0.3164	0.8999
Bus	0.0878	0.9923
Car	0.4083	0.8333
Motorcycle	0.0426	0.9982
Truck	0.1291	0.9833
Animal cart	-0.2028	0.9589
Sewing machine	0.3101	0.9039
Water pump	0.2773	0.9231
Tractor	0.0237	0.9994

Source: National Social Protection Monitoring Survey, 2012. (First round)

The wealth index was then used to group households into five quintiles, i.e., into five groups of equal size, from lowest to highest wealth index, taking into account the size of the household. Thus, the poorest quintile (q_w4=1) is comprised of the 20 per cent poorest individuals (not households) as per the wealth index, while those individuals in the richest quintiles (q_w4=5) comprise the 20 per cent richest.

After generating the wealth quintiles for all individuals in the balanced sample using the information from the first round, the wealth quintiles are imputed to all individuals in all rounds. (See Stata code in annex)

References

ORC Macro (2004). 'The DHS Wealth Index', *DHS Comparative Reports*, No. 6, August. Claverton, MA, ORC Macro.

Filmer, D. and L. Pritchett (2001). 'Estimating Wealth Effects without Expenditure data — or teas: An application to educational enrollments in states in India', *Demography*, 38(1): 115–132.

Annex. Stata code for the Wealth index

- Preparing the variables to enter the principal component analysis

```
tab p1_q_07, gen(floor_)
tab p1_q_11, gen(water_)
tab p1_q_10, gen(sanit_)
gen crowd1              = hhn/p1_q_04
gen a_electr1           = cond(p1_q_08_1 <= 4,1,0)
gen a_house1            = cond(p1_q_02 == 1,1,0)
gen a_land1             = cond(p7_q_2 == 1,1,0)
gen a_radio1            = cond(p7_q_15_a == 1,1,0)
gen a_tv1               = cond(p7_q_15_b == 1,1,0)
gen a_satTVreceiver1 = cond(p7_q_15_c == 1,1,0)
gen a_mobile1           = cond(p7_q_15_d == 1,1,0)
gen a_fixed1            = cond(p7_q_15_e == 1,1,0)
gen a_bed1              = cond(p7_q_15_f == 1,1,0)
gen a_refrig1           = cond(p7_q_15_g == 1,1,0)
gen a_fan1              = cond(p7_q_15_h == 1,1,0)
gen a_bike1             = cond(p7_q_15_j == 1,1,0)
gen a_bus1              = cond(p7_q_15_k == 1,1,0)
gen a_car1              = cond(p7_q_15_l == 1,1,0)
gen a_motorcycle1       = cond(p7_q_15_m == 1,1,0)
gen a_truck1            = cond(p7_q_15_n == 1,1,0)
gen a_cart1             = cond(p7_q_15_o == 1,1,0)
gen a_sewmachine1       = cond(p7_q_15_p == 1,1,0)
```

```
gen a_waterpump1        = cond(p7_q_15_q == 1,1,0)
gen a_tractor1          = cond(p7_q_15_s == 1,1,0)
```

- Running the model:

```
factor floor_1-floor_7 water_1-water_15 sanit_1-sanit_11 crowd1 a_electr1 a_house1 a_land1
a_radio1 a_tv1 a_satTVreceiver1 a_mobile1 a_fixed1 a_bed1 a_refrig1  a_fan1 a_bike1 a_bus1
a_car1 a_motorcycle1 a_truck1 a_cart1 a_sewmachine1 a_waterpump1  a_tractor1
[aw=adj_weight_w1ew2ew3ew4] if f2 == 1 & line_no == 1, pcf factors(1)
```

- Generating the wealth index

```
predict double wealth4 if f2 == 1
```

- Generating the wealth quintiles at the household level (if `hhdummy'==1), but taking into account the size of the household, by multiplying the weight by hhn. Also notice that as the weights are not integer numbers we need to round them.

```
tempvar hhdummy quint
by f2 f1 (line_no), sort: gen byte `hhdummy' = (_n == 1)
xtile `quint' = wealth4 [fw = round((10^5)*hhn*adj_weight_w1ew2ew3ew4, 1)] if f2 == 1 &
`hhdummy' == 1, nq(5)
```

- Imputing the household wealth quintile to each individual in the household

```
by f2 f1, sort: egen q_w4 = max(`quint') if f2 == 1
```

- Imputing the wealth quintile calculated in round 1 to all rounds.

```
xtset indkey f2
forvalues r = 2(1)4 {
        replace q_w4 = L`= `r' - 1'.q_w4 if f2 == `r'
}
xtset, clear
tempvar mq_w4
by f2 f1, sort: egen `mq_w4' = max(q_w4)
replace q_w4 = `mq_w4' if missing(q_w4) & !missing(adj_weight_w1ew2ew3ew4)
```

Distribution of the wealth quintiles in round 4.

```
. tab q_w4 if f2 == 4 [fw = round(adj_weight_w1ew2ew3ew4)]

       q_w4 |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |  4,381,185       20.02       20.02
          2 |  4,391,785       20.07       40.08
          3 |  4,361,451       19.93       60.01
          4 |  4,376,974       20.00       80.01
          5 |  4,376,227       19.99      100.00
------------+-----------------------------------
      Total | 21,887,622      100.00
```