Construction of Consumption Aggregates for the Ethiopia Socioeconomic Survey (Wave 1)

Introduction

The Ethiopian Socioeconomic Survey (ESS1) is a collaborative project between the Central Statistics Agency (CSA) of Ethiopia and the World Bank Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) project.[1] The first wave of data collection (ESS1) was conducted in 2011-2012 and covered rural and small town areas throughout Ethiopia. The ESS1 collects detailed information on a wide variety of topics allowing for a detailed analysis on many dimensions of wellbeing. One common measure of poverty and wellbeing in developing countries is aggregated household consumption expenditures. Construction of consumption aggregates can be a complicated process. For the convenience of data users, a consumption aggregate was calculated using the ESS1 data and is available for download[2]. This document provides a description of how the consumption aggregate was calculated.

As a part of the ESS1, information was collected on consumption of 25 food items consumed over the last 7 days, 11 basic household goods (matches, batteries soap, etc.) over the past month, and 12 other expenditures (men's clothing, linens, etc.) over the past 12 months. The consumption data was collected on the third visit to the household, conducted between January and March 2012. Based on these data, the consumption aggregate was calculated and adjusted based on adult equivalencies and regional price differences. The ESS1 data was fairly straightforward to work with, having high internal consistency, and relying on a very limited range of units to record food consumption. Only three main corrections / constructions were made from the publicly available data:

1. Winsorizing of per capita quantity consumed of each food item at the 97th percentile

2. Construction of local prices to value home consumption (median price at most local level having at least 10 price observations)

3. Winsorizing the value of per capita total food consumption at 2nd percentile

Tight control of the unit-item combinations used to record food quantities meant that there was little need for conversion factors. Interviewers were trained to work with respondents to back out the quantity of consumption in standard units (kilograms, grams, liters, etc.). Thus, we relied heavily on the interviewers to correctly record food consumption in the given units. The interviewer's supervisor reviewed the questionnaire for implausible values at branch headquarters.[3]

Preliminaries

Household size and a measure of adult-equivalency are constructed. The following scale factors are used for constructing adult-equivalency.

---

[1] For more information on the LSMS and the LSMS-ISA go to www.worldbank.org/lsms
[2] The consumption aggregate file is included in the full ESS1 data which can be downloaded here.
[3] Each supervisor managed three interviewers.

Age Range	Male	Female
<= 1 yr	0.33	0.33
> 1 yr and <= 2 yrs	0.46	0.46
> 2 yrs and <= 3 yrs	0.54	0.54
> 3 yrs and <= 5 yrs	0.62	0.62
> 5 yrs and <= 7 yrs	0.74	0.70
> 7 yrs and <= 10 yrs	0.84	0.72
> 10 yrs and <= 12 yrs	0.88	0.78
> 12 yrs and <= 14 yrs	0.96	0.84
> 14 yrs and <= 16 yrs	1.06	0.86
< 16 yrs and <= 18 yrs	1.14	0.86
> 18 yrs and <= 30 yrs	1.04	0.80
> 30 yrs and <= 60 yrs	1.00	0.82
> 60 yrs	0.84	0.74

In addition, we use a regional price index, based on the index created by the Ministry of Finance and Economic Development (MoFED) in their Household Consumption Expenditure (HCE) 2010-2011 report[4].

Region	Price Index
Tigray	1.034
Afar	1.021
Amhara	0.949
Oromia	0.981
Somalie	1.132
Benshagul	0.958
SNNP	0.906
Gambelia	1.056
Harari	1.227
Diredwa	1.245

I. Food Consumption

---

[4] The HCE analytical report can be found at http://catalog.ihsn.org/index.php/catalog/3123. The spatial price index is located in Annex III.

The food consumption data in the ESS1 covers 25 separate food items that are commonly consumed in Ethiopia, and asks whether "you or others in your household" consumed any of the item over the part one week. For each item that was consumed, the household was asked how much was consumed, how much of the consumption came from purchases, how much was spent on these purchases, how much of the consumption came from own production, and how much came from gifts and other sources. The data is in long format (household-item level) with observations for all households for all items. Note that a small number of households report not consuming any of the items (70 households, 1.8% of the sample). Since it is implausible for a household to have zero consumption, these households were excluded from calculation of the consumption aggregate. These households can be identified by the variable *no_cons* included in the consumption aggregate data file.

A. Standardize Units

While most food quantities were reported in standard units, a few food items did require conversion. For these cases, conversion factors were utilized. These conversions factors were collected during a market survey conducted prior to the third wave of the ESS. The conversion factors are included in the full ESS1 data download package (***Food_CF_Wave1.dta***). There were some cases where quantities reported in nonstandard units could not be converted. When this was the case, the household was excluded from the consumption aggregate calculation (67 households, 1.7% of the sample). These cases are identified in the consumption aggregate data file using the variable *no_conv*.

B. Quantities Consumed

Each observation was checked to see if the reported total quantity consumed equaled the sum of the reported quantities consumed from each source (purchases, own production, gifts) and in 98.8% of observations it did. Note that less than 1% of household-items were obtained from more than one source, i.e., some purchased and some consumed from home consumption. Next, per capita amounts consumed were calculated, and winsorized at the 98th percentile for non-zero consumption for each item. That is, for each item, using only the observations that report non-zero consumption of the item, the 98th percentile level of per capita consumption was constructed, and observations that reported higher levels of per capita consumption were replaced with the 98th percentile value.

Code actually used:

```
gen pc_quant_sum = quant_sum / hh_size
gen tag=.


forval i=1/25 {
      su pc_quant_sum if q00==`i' [aw=pw], d
      _pctile pc_quant_sum if q00==`i' & pc_quant_sum!=0 [aw=pw], nq(100)
      replace tag=1 if pc_quant_sum>`r(r98)' & pc_quant_sum!=. & q00==`i'
      replace pc_quant_sum =`r(r98)' if pc_quant_sum>`r(r98)' & pc_quant_sum!=. &
q00==`i'
      su pc_quant_sum if pc_quant_sum!=0 & q00==`i' [aw=pw], d
      }
```
C. Calculate Prices

We assign households a price for each food item (ph) so that we can estimate the value of

food consumption for food produced at home or given as gifts. To achieve this, we first calculate prices paid by households who purchased a given food item (i.e., amount paid / amount purchased in standard units); ph is then determined for a given as the median price at the lowest geographical unit for which we have at least 10 price observations[5]. There are 12 households per EA, so for the most frequently purchased items (salt, onions) EA prices can often be constructed. If there are less than 10 price observations for that item at the EA, the next level up is used. (The geographical levels used, in ascending order, are EA, Kebele, Woreda, zone and region and national).

```
egen price_national = median(price), by(q00)
gen ph = price_national
gen p0 = price_national
gen p_level = 0

egen price_region = median(price), by(q00 saq01)
egen obs_p_region = count(price), by(q00 saq01)
replace ph = price_region if obs_p_region >= 10
replace p_level = 1 if obs_p_region >= 10

egen price_zone = median(price), by(q00 saq01 saq02)
egen obs_p_zone = count(price), by(q00 saq01 saq02)
replace ph = price_zone if obs_p_zone >= 10 & price_zone<10*price_region &
price_zone>price_region/10
replace p_level = 2 if obs_p_zone >= 10

egen price_woreda = median(price), by(q00 saq01 saq02 saq03)
egen obs_p_woreda = count(price), by(q00 saq01 saq02 saq03)
replace ph = price_woreda if obs_p_woreda >= 10 & price_woreda<10*price_region &
price_woreda>price_region/10
replace p_level = 3 if obs_p_woreda >= 10

egen price_kebele = median(price), by(q00 saq01 saq02 saq03 saq06)
egen obs_p_kebele = count(price), by(q00 saq01 saq02 saq03 saq06)
replace ph = price_kebele if obs_p_kebele >= 10 & price_kebele<10*price_region &
price_kebele>price_region/10
replace p_level = 4 if obs_p_kebele >= 10


egen price_ea = median(price), by(q00 saq01 saq02 saq03 saq06 saq07)
egen obs_p_ea = count(price), by(q00 saq01 saq02 saq03 saq06 saq07)
replace ph = price_ea if obs_p_ea >= 10 & price_ea<10*price_region &
price_ea>price_region/10
replace p_level = 5 if obs_p_ea >= 10

tab q00 p_level
```

Note that this does not flag or exclude from the analysis any outlier prices, it assumes that once you are taking the median over 10 prices, the effect of outliers is minimized.

D. Value of Consumption

Two different formula are used to compute the value of consumption based on whether or not the total per capita consumption was winsorized. We use purchase value + price *(owned and gifted amount) for those items that were not windsorized, and price * total amount consumed (corrected) for those that were in order to ensure we capture the new winsorized amount. If we believe that the quantities winsorized are mistakes (in recording or data entry) there is no particular reason to think that the purchase value is also a mistake for items that are

---

[5] We also only replace with the more local price if the local price is not more than 10 times the regional price ad not less than 1/10 of the regional price.

consumed entirely from purchases.

We then construct Paasche, Laspeyres and Fisher spatial price indices for comparison with the official MoFED price index which was calculated from the Household Consumption Expenditure Survey (HCE) from 2010/11. The MoFED index is included in the consumption aggregate file (*price_index_hce*) but are not applied in calculation of the aggregates. It is left to the use to decide on whether to use this index or calculate another.

Finally, we annualize food consumption by scaling up food consumption expenditures by a factor of 52.

II. Non-Food Consumption and Education Expenditure

The survey also asked about expenditure over the past one month on 11 basic household goods (matches, batteries, soap etc.), and expenditure over the past 12 months on 12 other expenditure categories (men's clothing, linens etc.). No particular processing was done of this data, except to drop spending on taxes and levies, and to annualize spending. A few questions on education expenses were also asked -- the amount spent on school fees, and the amount spent on school books, uniforms, stationary etc. -- over the past 12 months, for every household member attending school. These were also included in the non-food items, as was the total value of scholarships or other school assistance from the government or other organizations.

III. Total Expenditure

Consumption expenditures from these three sources (food, non-food purchases, education expenses) were combined for each household. The bottom 2nd percentile of per capita food consumption was winsorized.

Nominal per adult equivalent consumption is then calculated. Real per adult equivalent consumption is not included in the consumption aggregate data set. As mentioned above, the MoFED spatial price index is included to allow the user to correct for spatial variation. The data user may also wish to correct for price variation over time (especially if using data from multiple waves of the ESS). There are several options for making this correction including using the official inflation figures from CSA[6] or using survey based indices. It is left to the user to decide what method to use.

IV. Variables Included in Data File

- *household_id:* household identifier
- *ea_id:* enumeration area identifier
- *saq01:* region code
- *rural:* rural vs small town indicator
- *pw*: household survey weight
- *adulteq:* adult equivalent household size
- *hh_size*: basic household size
- *no_conv*: indicator for cases where at least on consumption quantity could not be

---

[6] According to CSA, the official inflation between February 2012 and February 2014 (ESS2) was 21%.

converted to standard units
- *no_cons*: indicator for cases where a household did not report any food consumption
- *food_cons_ann*: total (nominal) annualized food consumption
- *nonfood_cons_ann:* total (nominal) annualized nonfood consumption
- *educ_cons_ann:* total (nominal) annual expenditure on education
- *total_cons_ann*: total (nominal) annual consumption expenditure
- *price_index_hce:* spatial price index from MoFED, calculated from HCE
- *nom_totcons_aeq*: nominal total annual per adult equivalent consumption expenditure
- *cons_quint*: (population weighted) consumption quintiles based on *nom_totcons_aeq*