Construction of the consumption aggregate based on the Nigeria Living Standards Survey, 2018-19

06/17/2020

I. Context

The analysis of poverty and inequality is important input to policy making, aimed at improving the welfare of the population. At the same time, the process and choices involved in computation of the welfare aggregate are not trivial. There are many issues that require adoption of assumptions, taken to ensure that estimations are robust and do not require constant revisions.

The main objective of this note is to document the decisions, methodology and steps, which were taken while constructing the consumption aggregate based on the Nigeria Living Standards Survey 2018-19 (2018/19 NLSS). The details of the 2018/19 NLSS are reported in the Basic Information Document published along with 2018/19 NLSS database, while this note is solely focused on the consumption aggregate.

II. Construction of the welfare aggregate

It is difficult to overstate the significance of properly constructed welfare aggregate, since consumption aggregate serves as a basis for estimating the poverty line(s) and the welfare indicators for the country. The 2018/19 NLSS collected all necessary data to calculate the required components of consumption aggregate. Construction of consumption aggregate is a complex process, which combines the well-established procedures as well as value judgments reflecting the country context. A complete discussion of theory and issues involved in construction of consumption aggregate can be found in Deaton and Zaidi (Deaton & Zaidi, 2002), while this section highlights the key steps and decisions.

The consumption aggregate is the monetary value of food and non-food goods and services consumed by the household. Corresponding to the available data in 2018/19 NLSS and practices adopted in the previous living standards surveys in Nigeria, the consumption aggregate consists of 6 primary components:

1. Food consumption, including purchased, self-produced and received as a gift food items;
2. Meals or expenditures on food outside of home;
3. Non-food section, which includes frequently purchased items; non-food goods and services purchased on a regular basis and non-food expenditures incurred on less regular intervals, e.g. on semi-annual or annual basis;
4. Education expenditures;
5. Health expenditures;
6. Dwelling or rent expenditures.

i. Food consumption expenditures and spending on meals

The nominal food consumption expenditures are calculated as the total value of consumed food products, obtained from various sources including from purchases, self-produced and received as gifts etc. The food consumption module of the NLSS questionnaire asks if household consumed a particular food item in the last 7-day period and if so, the physical amount consumed at the household level. Households do not report the monetary value of food consumed, therefore, to obtain/calculate the value of consumption, one needs to utilize the purchase part of the module to impute the unit values of each food item.

The food consumption module covers 16 general food categories (grains, starchy roots and tubers, cereals, baked food, etc.), and more than 160 individual food items, including drinks and alcohol. The module includes questions on the sources of the consumed food: from purchases, self- produced and from other sources, like gifts, barter, in-kind payments, etc.

The estimation of the value of the food consumption is undertaken in three steps: 1) convert the consumed and purchased food quantities from non-standard units of measurement into the standard ones (e.g. from *mudu* to *kilograms*); 2) calculate the unit values, as a ratio of expenditures or cost of purchased food item to the standard quantity in kilograms; 3) then the unit values are used to obtain the value of food consumption for each item.

Households in Nigeria and across regions within the country use multiplicity of non-standard measurement units (e.g. pieces, bunches, heaps, bags, paint rubber, cups etc.) for different food items, which might affect the accuracy of food consumption. Within the context of NLSS survey the NBS team undertook a challenging task of creating high-quality conversion factors for each non-standard unit of measurement and for each food item, disaggregated by the regions where possible. For a very few non-standard units of measurement and items which lacked the conversion rate, the decision was to apply the conversion of similar food items, e.g. if conversion for heap of nuts was missing then conversion factor for heap of rice was used. Except for few cases, no treatment of outliers in quantities was applied.

In the second step the unit values for each food item and each household were calculated as a ratio of cost of purchases of food item to the quantity purchased in kilograms in each month.

$$uv_{i,m}^{h} = \frac{expenditures_{i,m}^{h}}{quantity_{i,m}^{h}}$$

Where: *uv* stands for unit value; superscript *h* here and elsewhere in this note denotes the household; subscript *i* denotes specific item (here food item); *m* – specific month.

Given the importance of unit values in computing the food consumption, the efforts were made to identify and replace the outliers in unit values, i.e. values in log terms higher or lower than 2.5 standard deviation from the mean have been identified and removed.

$$outlier\left(uv_{i,m}^{h}\right) = abs\left[\frac{\ln\left(uv_{i,m}^{h}\right) - mean\left(ln\left(uv_{i,m}^{strata}\right)\right)}{std\left(ln\left(uv_{i,m}^{strata}\right)\right)}\right] > 2.5$$

If a household reported a purchase of the food item, then the unit value is available, and the unit values are taken at the household level. In cases when the unit value is an outlier or missing (i.e. self-produced or was received as a gift, etc.), i.e. not available, the unit values represent the median values for specific month at the lowest available strata level (i.e. hierarchical imputation procedure): first taken from the level of enumeration area, then at state, zone, and finally at country levels, in a given sequence:

$$uv_{i,m}^{h} = uv_{i,m}^{h}, \quad if\ uv_{i,m}^{h}\ is\ not\ missing\ or\ not\ outlier$$
$$uv_{i,m}^{h} = median\ (uv_{i,m}^{ea}), \quad if\ uv_{i,m}^{h}\ is\ missing\ or\ outlier$$
$$uv_{i,m}^{h} = median\ (uv_{i,m}^{state}), \quad if\ uv_{i,m}^{h}\ is\ missing\ or\ outlier$$
$$uv_{i,m}^{h} = median\ (uv_{i,m}^{zone}), \quad if\ uv_{i,m}^{h}\ is\ missing\ or\ outlier$$
$$uv_{i,m}^{h} = median\ (uv_{i,m}^{country}), \quad if\ uv_{i,m}^{h}\ is\ missing\ or\ outlier$$

In the third and final step, the unit values are multiplied by the quantity consumed (including from own-production, food received as a gift or for free, etc.). The food consumption aggregate for a household is then a sum of imputed values of food consumption, across all food items the household consumed in the last 7 days, which is then annualized.

$$x_{Food,m}^{h} = \sum_{i=1}^{I} quantity_{i,m}^{h,cons} * uv_{i,m}^{h} * (\frac{365}{7})$$

The procedure for calculating the expenditures on meals consumed outside of home is different and more straightforward. The households directly report the amount spent on food consumed outside of home e.g. in cafeteria, streets, café and restaurants, by categories, like breakfast, lunch and dinner or snacks or drinks. Therefore, the monetary value of meals outside of home is a sum of all expenditures on meals for each household.

$$x_{Meals,m}^{h} = \sum_{i=1}^{I} spending\ on\ meal\ category_{i,m}^{h} * (\frac{365}{7})$$

ii. Non-food expenditures

The non- food component of the consumption aggregate includes a heterogenous and wide-ranging group of expenditures on non-food items:

(1) Recreational and culture items: newspapers, magazines, cinemas, tobacco, etc.

(2) Energy: petroleum, kerosene, liquid gas, diesel, charcoal, electricity, etc.

(3) Clothing and footwear

(4) Personal services: repairs of appliances, etc.

(5) Household goods and appliances: furnishing, small electric appliances, etc.

The non-food items that were included in the consumption aggregate are: i) not "lumpy", e.g. purchase of a car or long-term durable goods; ii) not used as inputs in the production or as investment; iii) not acquired for the purpose of transfer to other households.

Given that expenses for non-food items that were included generally take place with different frequencies, the questionnaire asked the household to recall their expenditure on these items using different recall periods: 7 day, the last month and the last year. For the non-food part of the consumption aggregate, the challenge is to decide what needs to be excluded. To avoid bias in ranking of households, non-food expenditures exclude infrequent expenses like mortgages and ceremonial spending (funerals, weddings, dowries etc.).

Non-food expenditures are valued at the purchase or self-reported acquisition value. Depending on the reference period the expenditures are annualized by factor of (365/7) for 7-day recall items or (12/1) for one month recall items, except for expenditures on dwelling repairs, which were included without annualization.

$$x^h_{Non-food,m} = \sum_{i=1}^{I} purchase\ of\ nonfood\ item^h_{i,m} * (appropriate\ annualiztion\ factor)$$

iii. Housing expenditures/ rent

Housing cost is defined as implicit value or benefit that household receives from occupying a dwelling and not the expenditures on purchasing the dwelling itself. To measure monetary flow of housing services, the concept of rent is the appropriate choice. Since most households own the dwelling and thus do not explicitly pay the rent, a hedonic regression model is applied to estimate coefficients and predict the rent. The dependent variable is actual rent paid in log form, regressed on a set of housing variables/attributes like, location, number of rooms, material of roof, material of floor, material of wall, amenities/utilities (toilet, water sources, garbage collection, etc.):

$$ln(R^h) = a_0 + \sum_{n=1}^{N} f_n(X^h_n)\, b^h_n + e^h$$

Where R^h is the actual, observed rent;

X^h_n is observable housing attributes/characteristics (n).

The imputed rent is a predicted value of housing from regression, that was re-transformed into Naira terms from log form, using Duan-Smearing transformation method (Duan, 1983) and applied to the households that own their dwelling or do not pay rent for the dwelling they reside in. The challenge in estimating the rent is that rental market in some areas specifically in rural areas is quite thin. To overcome this, the hedonic model is applied at more aggregated strata levels, instead of state levels. Also given some

housing characteristics, the model results in extremely low rent values in rural areas. To avoid this the minimum value of rent is limited from below at the minimum of actual rent in a specific location/stratum.

iv. Education expenditures

Education expenditures is important component of the consumption aggregate. While spending on education could be viewed as long-term investment and thus be excluded from the aggregate, it is also true that education directly contribute to the household welfare. Since, education expenditures are regular across all groups of households and there is positive relationship between school spending and consumption deciles, education expenditures are included in the aggregate.

Expenditures on education includes all school related expenses from pre-school to tertiary education: school fees, uniform, textbooks, meals and lodging, transport, gifts to teachers and services to school, private tutoring and other expenses for education and schooling. Education expenses were recorded for past school/academic year and no annualization factor was applied. Since the survey was administered throughout the calendar year, some households reported past expenses of completed academic year and some reported expected expenses for upcoming academic year.

$$x^{h}_{Education,m} = \sum_{i=1}^{I} actual\ or\ expected\ school\ spending\ category^{h}_{i,m}$$

v. Health care expenditures

The health module of NLSS extensively covers questions, reflecting health care seeking behavior of households. Health expenditures are recorded on consultations, medicines, laboratory exams, hospitalization charges, transport and other out of pocket costs related to health care. Elsewhere the motivation for excluding the health-related expenditures from the welfare aggregate is related to consideration of health cost as a “regrettable necessity”. If a member of household falls ill and incurs medical expense this will increase total expenditures and therefore household’s level of welfare when in fact, the opposite may be the case. In similar vein, some components of health expenditure have the characteristic to be a reaction to a shock, for which extraordinary means are used.

However, the analyses of the data showed that health expenditures should be included in the consumption aggregate. First the prevalence of health insurance in Nigeria is negligible, below 0.5 percent of population. Thus, access to health care is welfare enhancing and affects ranking of households. Second, close to 80 percent of health cost is attributed to the purchase of drugs and medicine, which are not directly subsidized. Again, excluding these costs will negatively affect the welfare ranking. Finally, the health expenditures are prevalent across all income groups and the share of health cost in total household budget is significant. There is a positive correlation between healthcare spending and consumption deciles, pointing to significant elasticity of health expenditures with respect

to total expenditures. This clearly implies that health is central component of household welfare, as well as important area of public policy.

$$x^{h}_{Health,m} = \sum_{i=1}^{I} spending\ on\ healthcare\ category^{h}_{i,m}$$

vi. Adjustments to consumption aggregate

Once the components have been calculated and added up, there are two important corrections that need to be applied: 1) adjustment for regional and temporal price differences; and 2) adjustment for household composition.

a) Price adjustment across regions and months

Nominal expenditures are affected by substantial price differences between urban and rural areas, and between different parts of the country. In addition, there are temporal price differences whereby prices faced by households at the beginning of the survey might differ from prices at the end of the survey timeline. To ensure comparability of the aggregate across geographical areas and regional price patterns, a deflator needs to be calculated and applied. Since regional price indices were not available, this adjustment was undertaken using information on the budget shares of food items as well as the implicit prices or unit values of food items, obtained directly from the survey.

The price deflator was obtained by constructing a Paasche price index, while alternative was Laspeyres index. The difference between the two is how the relative prices are weighted. In the case of the former, the relative prices are weighted by the budget share of the item for the household itself, i.e., if a household spends a large share of its total budget on a certain food item, it receives a higher weight. In contrast, the Laspeyres formulation weights the relative prices according to the budget share of that item for a hypothetical, average reference household. In this case, all prices faced by households, irrespective of their consumption of an item, receive the same weight as that of the reference household.

The implications of the choice of price index are important, if there are significant differences in relative prices faced across regions within a country and if the hypothetical reference household is not representative of significant parts of the country. If this is the case, Paasche allows the most flexibility to incorporate these differences in relative prices and budget shares. In addition, construction of Laspeyres requires that composition of the price basket is largely similar across regions, but this is rarely the case as regions consume differential diets. Based on this it was decided to use the Paasche formulation for the spatial and temporal price index. Unlike other candidate measures, Paasche incorporates the quantities of each item consumed by the household in a specific month and weights them accordingly.

The Paasche price index for each household is obtained with the following formula:

$$P_m^h = \frac{\sum_i uv_{i,m}^h * quantity_{i,m}^h}{\sum_i uv_i^0 * quantity_i^h}$$

Where uv_i^0 is the price of food item *i* for the reference strata 0, which is in our case a median for total population, covering the survey period, i.e. at country level.

An intuitive interpretation of the Paasche index is that it represents the ratio between the cost of the food basket spent by a household, in a specific month over the cost of the same basket for the average, national, hypothetical household, here indexed by "0". The Paasche index can also be presented in another way (Deaton & Zaidi, 2002):

$$P_m^h = \left(\sum_{i=1}^{I} w_{i,m}^h * \frac{uv_i^0}{uv_{i.m}^h} \right)^{-1}$$

Where $w_{i,m}^h$ is the budget share of food item *i* in food basket for a given household in a specific month

$\frac{uv_{i.m}^0}{uv_i^h}$ is the relative price of the item *i*.

The Paasche index was first calculated at the level of each household. Median national prices for each item, calculated by applying household weights, were used as reference prices. Then, to stabilize the index, the median value of household price indices was taken at state level. A strata/state level index was preferred to a household Paasche index, to avoid the effect of influential outliers in some households, as well as of cases in which the household food consumption basket heavily concentrated on a few food items or most of its budget spent eating outside of home.

$PI_m^{state} = meadin(P_m^h), over\ state/strata$

b) Adjustment for household composition

Since, the purpose of consumption aggregate is to capture the welfare measure at the individual rather than at the household level, the total expenditure is adjusted by household composition based on household demographic variables. The resulting welfare measure can then be attributed to each household member rather than to the household unit. To this end, the welfare aggregate in addition to spatial/temporal deflation is adjusted using household size, i.e. transformed into per-capita terms, which assumes that all individuals in the household have the same needs and that consumption is shared equally among household members. An alternative is to calculate and apply adult equivalent household size, whereby an attempt is made to capture the differences in consumption practices/requirements by age and sex. The key problem with adult equivalent scale is its arbitrariness,

related to the difficulty of choosing an appropriate scale. In contrast, using the household size is more appealing due to its simplicity and transparency of assumptions.

$$size_m^h = \sum_{i=1}^{I} memeber\ of\ a\ household_{i,m}^h$$

vii. Putting all components together

Once all components of aggregate are added up for each household, the resulting nominal consumption aggregate needs to be adjusted for household size and spatial and inter-temporal price differences, observed across regions and months of survey:

$$Consagg_{adj}^{pc} = \frac{x_{Food,m}^h + x_{Meals,m}^h + x_{Non-food,m}^h + x_{Education,m}^h + x_{Health,m}^h + x_{Housing,m}^h}{PI_m^{state} * size_m^h}$$

Where:
$Consagg_{adj}^{pc}$ is spatially and temporally adjusted consumption aggregate in per capita terms
$x_{Food,m}^h$ is food consumption expenditures for specific household in specific month;
$x_{Meals,m}^h$ is expenditures on food consumed outside of home for specific household in specific month;
$x_{Non-food,m}^h$ is non- food expenditures for specific household in specific month;
$x_{Education,m}^h$ is expenditures on education for specific household in specific month;
$x_{Health,m}^h$ is expenditures on health care for specific household in specific month;
$x_{Housing,m}^h$ is housing expenditures, rent for specific household in specific month;
PI_m^{state} is spatial and temporal price deflator, for each state in specific month;
$size_m^h$ is the household size.

III. Data file

The estimates of consumption aggregate as per procedure outlined above is presented in the Stata file "***totcons.dta***". The file contains a set of the variables, as follows:

Table 1: Variables contained in the totcons.dta file

Variable	Description	Type
hhid	Household identifier	
sector	Rural-urban location	Geographic identifiers
zone	Geo zone location	
state	State location	
lga	LGA location	
ea	Enumeration area	
month_of_interview	Month of interview date	Time period
year_of_interview	Year of interview date	
hhsize	Total number of household members	
wt_final	Sampling weight (household), final	Statistical weights
popw	Population weight, final (= wt_final * hhsize)	
food_own#	Consumption expenditures for food group # from self-production or received as a gift	Annual household expenditures by category
food_purch#	Consumption expenditures for food group #, purchased	
food_meals20	Expenditures on food consumed outside of home	
nonfood21	Expenditures on transport and other fares	
nonfood22	Expenditures on fuel	
nonfood23	Expenditures on electricity	
nonfood24	Expenditures on entertainment	
nonfood25	Expenditures on personal services	
nonfood26	Expenditures on household nonfood items	
nonfood27	Expenditures on clothing, footwear	
nonfood28	Expenditures on petrol	
edu29	Expenditures on school tuition	
edu30	Expenditures on other education related goods and services	
health31	Expenditures on healthcare consultations and drugs	
health32	Expenditures on hospitalization	
rent33	Rent/housing expenditures (imputed for owners and actual for renters)	
reg_def	Spatial and temporal price deflator	
totcons_pc	Total consumption expenditures, annual, nominal, in per capita terms	
totcons_adj	Total consumption expenditures, annual, adjusted for spatial deflation, in per capita terms	Final welfare aggregate

It is important to note that data file contains observations for Borno state. As it is mentioned in the Basic Information Document, the sample from Borno state was not representative of the whole state since only households from “accessible” areas were interviewed. The security situation in Borno was unstable at the time of the survey, preventing the enumerators from safely visiting all randomly selected households. Thus, the Borno sample is considered statistically non-representative and not comparable to other states. Still the decision was taken to keep the Borno state observations in the data file, however the users are advised against comparing the Borno sample to other states, in other words the Borno sample must be treated separately from the data for other states.