The accuracy of statistical samples:
How Egyptian society is depicted is income, expenditure, and consumption research

Dina Abdallah

Inaccuracy for security reasons:

Income and expenditure date were collected starting 1957 by the Central Authority for Public Mobilization and Statistics (CAPMAS) on irregular basis then was collected every five years between 1990-1991 and 2008-2009 and every two years in 2011, 2013, and 2015. Income and expenditure statements include a huge amount of raw data on income distribution, expenditure patterns and how both as well as poverty rates differ according to geographical location, family size, the nature of jobs, and other social characteristics of individuals and families. Therefore, those statements are a very rich source of information that can be used to study poverty, inequality, and inflation. Despite the amount of data that is collected every two years and for security reasons, CAMPAS only makes 50% of the sample data it collected after 1999 available. The available data will be referred to in this paper as the "partial sample."

Most of the studies that examine poverty and inequality in Egypt depend in their measurements and analyses on the partial sample based on the assumption that it represents the complete sample and that the data from both are extremely close as asserted by CAMPAS. Despite the fact that the differences between complete and partial samples in most

years are not remarkable, as will be demonstrated in tables (1) and (2), these seemingly slight differences might affect the accuracy of the results that aim at tracing the development of prices or poverty and hunger lines. That is why the following analysis does not only aim at underlining the differences between partial and complete samples, but also attempts to devise a correction coefficient based on the difference detected between the two samples then using this coefficient to arrive at more accurate numbers and percentages of the hungry based on the alternative poverty line proposed in Mohamed Sultan's paper "A clearer vision of the rock-bottom."

In an attempt to identify the difference between the partial and complete samples in the data of year 2015 only, it becomes clear that all indicators calculated from the partial sample using the same methodology applied by CAMPAS are different with a variable coefficient and range between 3% and 36% as demonstrated in Table (1). Based on this, it is only possible to obtain more accurate indicators if the complete sample is made available.

Table (1): Differences between poverty rates in complete and partial samples for year 2015:

Indicator	Partial sample	Complete sample	Correction coefficient
Percentage of individuals below hunger line	7.2٪	5.30%	1.36
Percentage of individuals below poverty line	23.05٪	27.80%	0.83
Percentage of poverty in families of 10 or more members	73٪	75%	0.97
Percentage of poverty in families between 8 and 9 members	67.70٪	65%	1.04
Percentage of poverty in families between 6 and 7 members	45.70٪	44%	1.04
Percentage of poverty in families between 4 and 5 members	20.70٪	20%	1.04
Percentage of poverty in families between 1 and 3 members	5.30٪	6%	0.88

If the focus in the difference between the partial and complete samples is only on the coefficient subject of research, the line of hunger, it will be obvious that it differs throughout the year as demonstrated in Table (2).

Table (2): Differences between percentage of the hungry in partial and complete samples (2010-2015):

Year	2010- 2011	2012- 2013	2015
Percentage of individuals below hunger line in complete sample	4.8%	4.4%	5.3%
Percentage of individuals below hunger line in partial sample	%4	%3.6	%7.2
Correction coefficient	1.2	1.2	0.74

Based on identified differences between the two samples, it is possible to propose correction coefficients of 1.2, 1.2. and 0.74 for years 2010-2011, 2012-2013, and 2015, respectively.

Table (3) demonstrates the percentages of individuals under the official hunger line compared to those proposed in Sultan's paper. For more accuracy in the proposed results, correction coefficients were applied to the alternative percentage of the hungry, hence leading the hunger percentage to appear as a range, as demonstrated in the last row of the table.

Table (3): Differences between individual under official and alternative hunger lines (2010-2015):

Year	2010- 2011	2012- 2013	2015
Percentage of individuals under official hunger line	4.8%	4.4%	5.3%
Percentage of individuals under alternative hunger line	%9.6	%9.3	%9.6
Percentage of individuals under hunger line after applying correction coefficient	9.6%-11.52%	9.3%-11.16%	7.29%-9.6%

Although the methodology proposed in Sultan's paper would yield percentages of poverty and hunger that are closer to reality than those stated in official data, even the numbers resulting from the alternative calculations are still less than the actual ones and even after applying the correction coefficient. This is attributed to the nature of the official complete sample itself because studies on poverty

and inequality are affected by how far the complete sample represents the Egyptian society. Below is a detailed account of what is meant by the nature of the sample.

Too rich to be held accountable, too poor to be visible:

Many economic studies tackle the accuracy of Egyptian official statistics on high-income shares .[52] These studies provide statistical evidence that high-income shares in Egypt, whether the highest 10% or 1%, are much less than the real numbers on the ground, which eventually affects inequality estimates. The contrast between inequality estimates and reality becomes obvious when official statistics of inequality come out similar to those of Scandinavian countries in the 1980s. This contradiction between statistics and reality drives researchers to look into the accuracy of income distribution data for the richer segments of society and to propose other statistical means to adjust official calculations and make them as close as possible to reality[53]. Yet regarding low-income shares, the question is whether official

52- Johan A. Mistiaen and Martin Ravallion (2003). *Survey compliance and the distribution of income*.

Anton Korinek (2006, 2007). *Excessive Dollar Borrowing in Emerging Markets Balance Sheet Effects and Macroeconomic Externalities*.

53- Anthony B. Atkinson, Thomas Piketty, and Emmanuel Saez (2011). *Top Incomes in the Long Run of History*.

Frank Cowell and Victoria-Feser (1996a and 1996b). *Poverty measurement with contaminated data: A robust approach.*

data accurately reflect the status of the poor and the hungry and whether they are thoroughly represented in the income and expenditure statement. Several studies argue that the answer is in the negative, possibly owing to the difficulty of conducting interviews with the poor because they live in remote areas or illegally across cities and in some cases have no permanent residence, which makes it more likely that their names are not on the lists of municipal authorities. Also surveys about the poor always lack information about one particular group: the homeless.

Table (4) shows that the homeless are not listed in housing surveys as well. Residents of slum areas are represented by a very small percentage (0.02%) of the total 2015 sample while the percentage of individuals living in poor areas in the same year is allegedly 2%. This example can explain why the percentage of the hungry in 2015 according to the alternative methodology appears less than that of previous years despite the fact that the rise of the hunger line was the highest amongst the years subject of the study.

Table (4): Types of Housing:

Type of housing	Number	Percentage
Country house	2085	17.39

Type of housing	Number	Percentage
Villa	22	0.18
Apartment	9125	76.12
More than one apartment	185	1.54
One or more separate room	148	1.23
One or more room in a housing unit	421	3.51
Tent, hut, cave, slum areas… etc.	2	0.02
Total	11988	100

This lack of representation fails to reflect the fact that even among the poor there are difference income levels as there is a segment that lies right under the poverty line and another that suffers from extreme poverty and is not documented in the survey. Some policy makers who attempt to effect real change might focus on channeling resources towards those closer to the poverty line, which means others below them might be overlooked. That is why it is necessary to include other calculations such as poverty gaps, meaning how far the poor are from the poverty line, in order to overcome this problem.

Income or expenditure data?

When calculating poverty indicators, CAMPAS mainly relies on expenditure rather income data even though there is also a discrepancy between the income and expenditure of an individual or a family, referred to in economics as the saving rate. This means that if the income decreases by 10%, consumption will most likely not decrease by the same percentage but rather by a lesser one, especially among low-income segments of society. However, in developing countries such as Egypt, analysts prefer using expenditure data as an indication of living standards for the following reasons:

1- In the short term, expenditure data reflect more accurately the resources owned by a family.

2- In the long term, expenditure data provide information on income on other dates both in the past and the future.

3- In poor countries, it is difficult to accurately measure incomes because of its multiple sources and the integration of large numbers into the informal sector.

Figure (10) -Annexes

However, consumption can for many reasons be a misleading indicator of welfare even after introducing modifications to consumption indicators. This is because poorer

families have lesser opportunities at saving or getting loans, which is demonstrated below by the graph that compares the total income with the total expenditure for each 20% of income levels.

The graph shows that the gap between income and consumption, which is the ability to save, only appears at the beginning of the third income level (LE 30,000 per year and more). This makes the well-to-do more capable of taking income shocks and changes in their expenditure patterns or more capable to avoid particular inflation rates through changing their expenditure patterns. The graph also underlines the fact the for the poor the income is exactly the same as the expenditure. As for income levels that are right above the poverty line, which means they are threatened with poverty, their current income might enable them to save and expenditure data alone will not be enough to examine how their expenditure patterns would respond to any changes in their income levels. This necessitates the availability of income data in order to make it possible to study the behavioral patterns of segments of society that are threatened with poverty.

The individual versus the household:

Household surveys in their conventional forms do not allow

for direct measurement of the income and consumption of each individual. In most cases, the head of the household is interviewed about the entire family and after this, per capita calculations are made. The results are more harmonious than they are in reality. Income and consumption data collected in this manner do not reflect inequality within each family since they are based on the assumption that members of the family are paid and spend the same amounts. The result is misleading conclusions that do not reflect the reality of inequality and poverty within families. In fact, one of the studies dealing with this issue revealed that relying on income and consumption data per household only to measure poverty and inequality can reduce actual percentages of both by more than 25%.

Subjective perception of poverty versus objective poverty:

The poverty indicator calculated based on the Household Income, Expenditure, and Consumption Survey (HIECS) is incapable of underlining the subjective perception of poverty, meaning poverty from the point of view of the poor. It is extremely important to know the effect of policies on the way the poor see themselves compared to how other segments of society see them. It is possible to use the Af-

ro-parameter survey for this purpose. For example, in the sixth round for year 2015, Afro-parameter stated that 20% of the survey sample said their conditions are worse or much worse than others. On the other hand, the percentage of objective poverty as calculated by HIECS for 2015 was 23.5%, which shows the difference between subjective estimates and objective calculations that rely on conventional statistics. This difference is not necessarily in favor of conventional statistics as is the case in the previous example. In most cases, subjective poverty is measured through asking individuals about an amount of money that they consider necessary to satisfy their basic needs. The subjective line of poverty can be determined based on the deduced average of an individual's share of those basic needs.

Based on the above, it is possible to reach a number of recommendations that render the process of measuring poverty and hunger more accurate through the following:

- Having access to the complete sample collected by CAMPAS in order to reach more accurate numbers through using HIECS
- Making available more accurate data on income together with consumption in order to obtain more accurate information on the poor

- More representation of the homeless poor as well as of the well-to-do in order to have more accurate measurements of poverty and inequality
- Placing more emphasis on individual income and consumption
- Measuring subjective perception of poverty to estimate the efficiency of government policies that target the poor