Food and Agriculture Organization of the United Nations

FAO Statistics Division

Working Paper Series

ESS / 14-05

REFINEMENTS TO THE FAO METHODOLOGY FOR ESTIMATING THE PREVALENCE OF UNDERNOURISHMENT INDICATOR

September 2014

ESS Working Paper No. 14-05, September 2014

REFINEMENTS TO THE FAO METHODOLOGY FOR ESTIMATING THE PREVALENCE OF UNDERNOURISHMENT INDICATOR

Nathan Wanner
Carlo Cafiero
Nathalie Troubat
Piero Conforti

Food and Agriculture Organization of the United Nations
Rome, 2014

Refinements to the FAO Methodology for estimating the Prevalence of Undernourishment Indicator

Nathan Wanner, Carlo Cafiero, Nathalie Troubat, Piero Conforti

Statistics Division

Food and Agriculture Organization of the United Nations

I. Introduction

The FAO prevalence of undernourishment (PoU) indicator monitors progress towards Millennium Development Goal target 1C of halving, between 1990 and 2015, the *proportion* of people suffering from hunger [1]. Estimates of the number of undernourished (NoU) - calculated by multiplying the PoU by the size of the reference population - are used to monitor progress towards the World Food Summit goal of reducing by half the *number* of people suffering from undernourishment [2]. The PoU indicator is defined as the probability that a randomly selected individual from the reference population is found to consume less than his/her calorie requirement for an active and healthy life. It is written as:

$$PoU = \int_{x<MDER} f(x)dx$$

where *f(x)* is the probability density function of per capita calorie consumption.

The parameters needed for the calculation of the indicator are: the mean level of dietary energy consumption (DEC); a cut-off point defined as the Minimum Dietary Energy Requirement (MDER); the coefficient of variation (CV) as a parameter accounting for inequality in food consumption; and a skewness (SK) parameter accounting for asymmetry in the distribution. The DEC as well as the MDER are updated annually, with the former calculated from the FAO Food Balance Sheets. The MDER is calculated as a weighted average of energy requirements according to sex and age class, and is updated each year from UN population ratio data. The inequality in food consumption parameters are derived from National Household Survey[1] data when such data is available and reliable. Due to the limited number of available household surveys, the inequality in food access parameters are updated much less frequently over time than the DEC and MDER parameters.

To implement this methodology it is necessary to: (i) choose a functional form for the distribution of food consumption *f(x)*; (ii) identify values for the three parameters, that is, for mean food consumption (DEC), its variability (CV) and its asymmetry (SK); and (iii) compute the MDER threshold. The probability density function used to infer the habitual levels of dietary energy consumption in a population, *f(x)*, refers to a typical level of daily energy consumption during a year. As such, *f(x)* does not reflect the possible implications of insufficient food consumption levels that may prevail over shorter periods. Both the

[1] National household surveys include household income and expenditure surveys (HIES), household budget surveys (HBS) and living standard measurement studies (LSMS).

probability distribution *f(x)* and the MDER threshold are associated with a representative individual of the population, of average age, sex, stature and physical activity level.

This paper will first discuss refinements to the choice of the probability density function for the calculation of the PoU with a data-driven criterion for the selection of the functional form. Descriptions of the first two parameters needed for the calculation of the PoU, the DEC and MDER, will be given in the next two sections. The later sections will present revised methods for estimating the CV and SK parameters both directly (when household survey data is available), and indirectly using a regression (in the absence of reliable household survey data). Lastly, a discussion of the limitations of the methodology will be given before concluding.

II. Functional Form

The FAO methodology for the calculation of the prevalence of undernourishment uses a probability framework in which the distribution of per capita calorie consumption of the representative individual is characterized. The use of such a framework is necessary, as data typically are not available on individual food consumption and requirements, but rather for household acquisition. Starting with the estimates of undernourishment produced for the Sixth World Food Survey in 1996, the distribution was assumed to be lognormal. This model is very convenient for the purposes of analysis, but has limited flexibility, especially in capturing the skewness of the distribution.

As part of the revisions made for the 2012 edition of *The State of Food Insecurity in the World Report*, the methodology moved away from the exclusive use of the two parameter lognormal distribution to adopt the more flexible three parameter skew-normal and skew-lognormal families [3]. In the case of the lognormal distribution, the skewness can be written as function of the CV as:

$$SK = (CV^2+3)*CV \quad (1)$$

This implies that the SK for the lognormal distribution is completely determined by the CV derived from household survey data. The flexibility gained from the additional parameter allows for independent characterization of the asymmetry of the distribution.

The skew-normal distribution can be considered a generalization of the normal distribution that can account for departures from normality to a certain degree, corresponding to skewness values within the approximate range (-0.995, 0.995) [4]. The distribution cannot be evaluated at higher levels of asymmetry, and so ways to deal with higher degrees of skewness need to be found. One solution is to consider only the restricted range of the skewed-normal distribution in the calculation of the PoU. Another solution is to add another level of flexibility in which the functional form for the distribution itself is allowed to change, based on the level of asymmetry in the data. The identification of the appropriate combination of functional forms as well as the level of asymmetry at which to change functional forms motivates the investigations below.

The simplest way to handle skewness outside of the range of the skewed-normal distribution is to place a ceiling on the SK parameter (such as 0.99) and to use this limit for higher degrees of asymmetry. Figure 1 shows the implementation of this approach (referred to as **Function 1**) – in (a) the PoU is shown as a function of the SK parameter with the other parameters fixed (DEC equal to 2000, MDER equal to 1800, and CV equal to 0.35) and in (b) the density function is shown the with the same parameters fixed but with the SK equal to zero (corresponding to the normal distribution), 0.75, and 0.99 (the ceiling). High levels of asymmetry in the data may indicate that the skew-normal distribution is not the appropriate model, and alternative criterions for the selection of the functional form are described below.

As a first alternative to the application of the skewed-normal distribution described above, consider replacing the ceiling with a new value W, and evaluating the log-normal distribution for skewness values higher than W. If we denote the PoU evaluated using the lognormal distribution as PoU_{LN}, we can write this criterion for the choice for the distribution (**Function 2**) as:

$$PoU = PoU_{LN}(DEC, CV, SK, MDER), \qquad SK \geq W \qquad (2a)$$

$$PoU = PoU_{SN}(DEC, CV, SK, MDER), \qquad SK < W \qquad (2b)$$

Although the two different functional forms for the distribution do allow for a wider range of levels of asymmetry to be captured, discontinuities in the PoU occur as the functional form transitions from one to the other. An intermediate distribution may help to link such a gap, and this is the motivation behind the criterion below for the choice of the functional form.

As a modification of the criterion described above, consider using the log-skewed-normal distribution[2] (denoted by PoU_{LSN}) as an intermediate between the transition of the functional form from the skewed-normal to the log-normal, as written below:

$$PoU = PoU_{LN}(DEC, CV, SK, MDER), \qquad SK \geq (CV^2 + 3)CV \qquad (3a)$$

$$PoU = PoU_{LSN}(DEC, CV, SK, MDER), \qquad W < SK < (CV^2 + 3)CV \qquad (3b)$$

$$PoU = PoU_{SN}(DEC, CV, SK, MDER), \qquad SK \leq W \qquad (3c)$$

In the criterion written above (**Function 3**), the skewness implied theoretically by the lognormal is used both as a floor for the application of the lognormal and as a ceiling for the application of the log-skewed-normal. The fixed switch point W is used as a floor for the application of the log-skewed-normal and as a ceiling for the application of the skewed-normal[3].

Figure 2 shows how the pdf changes for this function with SK values of 0.25, 0.75 and 1.5, and Figure 3 shows the PoU for Function 3 and a switch point of 0.4 in 3 dimensions using the same color legend. The increased flexibility, both in terms of an additional parameter and in terms of the choice of the functional form allowing for a smooth transition, has led to the selection of Function 3 for the calculation of the PoU included in the *The State of Food Insecurity in the World 2014*. The resulting model allows for improvements in inequality in food access to be accounted for, such as those made by food intervention programs targeting specific subpopulations, permitting the smooth transition all the way to a distribution in which there is symmetric access to food. In the next section, we will give an overview of the DEC parameter and how it is projected.

III. Estimating and projecting mean food consumption

To compute per capita DEC in a country, FAO has traditionally relied on Food Balance Sheets, which are available for more than 180 countries. This choice was due mainly to the lack, in most countries, of suitable surveys conducted regularly. Through data on production, trade and utilization of food commodities, the total amount of dietary energy available for

[2] For an application of the log-skewed-normal distribution, see [5].

[3] The transition skewness value that minimizes the change in the PoU as the model moves to the skew-normal was determined by simulation studies and has been set to 0.75 for Function 2 and .4 for Function 3.

human consumption in a country for a one-year period is derived using food composition data, allowing computation of an estimate of per capita dietary energy supply.

During the revision for *The State of Food Insecurity in the World 2012* a parameter that captures food losses during distribution at the retail level was introduced in an attempt to obtain more accurate values of per capita consumption. Region-specific calorie losses were estimated from data provided in a recent FAO study [6] and ranged from 2 percent of the quantity distributed for dry grains, to 10 percent for perishable products such as fresh fruits and vegetables.

The latest data from food balance sheets refer to 2011; therefore, additional sources were needed to estimate the DEC for the last three years, 2012–2014. The main source for 2012 and 2013 estimates was projections prepared by the Trade and Market Division of FAO. The Holt-Winters distributed lag model was used to project the DEC for 2014; in some cases, this model was also applied to compute projections for 2012 and 2013, when data from the Trade and Market Division were not available or unreliable. The Holt-Winters model uses a process known as exponential smoothing, which attributes higher weights to more recent data and progressively less weight to older observations. Weights decrease in each period by a constant amount, which lies on an exponential curve. For countries showing peculiar patterns, other simpler forecasting models were used, such as linear or exponential trends.

IV. Estimating the MDER threshold

To calculate the MDER threshold, FAO employs normative energy requirement standards from a joint FAO/WHO/United Nations University expert consultation in 2001 [7]. These standards are obtained by calculating the needs for basic metabolism – that is, the energy expended by the human body in a state of rest – and multiplying them by a factor that takes into account physical activity, referred to as the physical activity level (PAL) index.

As individual metabolic efficiency and physical activity levels vary within population groups of the same age and sex, energy requirements are expressed as ranges for such groups. To derive the MDER threshold, the minimum of each range for adults and adolescents is specified on the basis of the distribution of ideal body weights and the mid-point of the values of the PAL index associated with a sedentary lifestyle (1.55). The lowest body weight for a given height that is compatible with good health is estimated from the fifth percentile of the distribution of body mass indices in healthy populations [7].

Once the minimum requirement for each sex-age group has been established, the population-level MDER threshold is obtained as a weighted average, considering the relative frequency of individuals in each group as weights. The threshold is determined with reference to light physical activity, normally associated with a sedentary lifestyle. However, this does not negate the fact that the population also includes individuals engaged in moderate and intense physical activity. It is just one way of avoiding the overestimation of food inadequacy when only food consumption levels are observed that cannot be individually matched to the varying requirements.

A frequent misconception when assessing food inadequacy based on observed food consumption data is to refer to the mid-point in the overall range of requirements as a threshold for identifying inadequate energy consumption in the population. This would lead to significantly biased estimates: even in groups composed of only well-nourished people, roughly half of these individuals will have intake levels below mean requirements, as the group will include people engaged in low physical activity. Using the mean requirement as a threshold would certainly produce an overestimate, as all adequately nourished individuals with less than average requirements would be misclassified as undernourished [3].

FAO updates the MDER thresholds every two years based on regular revisions of the population assessments of the United Nations Population Division and data on population heights from various sources, most notably the Monitoring and Evaluation to Assess and Use Results of Demographic and Health Surveys project coordinated by the United States Agency for International Development (USAID). This edition of *The State of Food Insecurity in the World* uses updated population estimates from the 2012 revision published by the United Nations Population Division in June 2013. When data on population heights are not available, reference is made either to data on heights from countries where similar ethnicities prevail, or to models that use partial information to estimate heights for various sex and age classes.

V.Improved Procedures for estimating the Coefficients of Variation

As mentioned, Variability (CV) and skewness (SK) are derived from NHS data when they are available. NHS are typically designed to collect data used for poverty analysis and to update the composition of the commodities basket used to compile consumer price indices. As these surveys typically collect information on food as part of the expenditure module, they are considered a readily available source of information that can be used to conduct food security analyses and from which to derive variability (CV) and skewness (SK) parameters.

i. Treatment Methods for Data from National Household Surveys

When the data from NHS surveys are taken as observations of individual habitual consumption, they are inherently very noisy, i.e. characterized by a high degree of unexplained variability. As such, it is essential to apply some sort of data treatment method before the estimation of the inequality in food access parameters[4]. For those surveys that are analyzed in partnership with National Statistical Offices, a range of different data treatment methods are sometimes applied, in agreement with country representatives. For those data for which data treatment is not applied by a National statistical Office, the different methods presented here were investigated to find a unified approach for data treatment.

The first data treatment method examined here (this method will later be referred to as **Method 1**) is the use of the well-established interquartile range (IQR). After defining Q1 as the 25th percentile in our data and Q3 as the 75th percentile, the IQR can be written as [Q3 – Q1]. Extreme values may then be identified as values that lie outside of the range [Q1 – D x IQR, Q3 + D x IQR] where D is a modifiable distance parameter specifying how strict the outlier detection method is. Using 2 as the value for the distance parameter, values of per