Statistics Division

Working Paper Series

NO: ESS /

$$\pi_n = \sum_{k=1}^{K} \left(\frac{y_n^k}{y_n}\right)\left(\frac{p_n^k}{p^k}\right)$$

FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

منظمة الأغذية والزراعة للأمم المتحدة | 联合国粮食及农业组织 | Food and Agriculture Organization of the United Nations | Organisation des Nations Unies pour l'alimentation et l'agriculture | Organización de las Naciones Unidas para la Agricultura y la Alimentación

THE FAO PARAMETRIC VERSUS THE IFPRI NON-PARAMETRIC APPROACH TO ESTIMATING THE PREVALENCE OF UNDERNOURISHMENT: *Issues Relating to the Use of Household Level Data from National Household Surveys*

Sibrián, Ricardo[1]

Naiken, Loganaden[2]

Mernies, Jorge[3]

1/ Senior Statistician, Global Statistics Service (ESSG), Statistics Division, FAO

2/ Former Chief, ESSA, Statistics Division, FAO

3/ Former Chief, ESSA, Statistics Division, FAO

TABLE OF CONTENTS

I. INTRODUCTION **3**

II. THE FAO PARAMETRIC APPROACH **4**

a. Derivation of the Formula for pU *4*

b. Evaluation of the Formula for pU *7*

c. Estimation of r_L *8*

d. Estimation of μ_X *10*

e. Estimation of CV_X *11*

III. THE IFPRI NON-PARAMETRIC APPROACH **12**

IV. THE INAPPROPRIATENESS OF THE NON-PARAMETRIC APPROACH **13**

a. The Basic Methodological Framework *13*

b. Reliance on Household Level Data that are subject to Biases and Errors *14*

V. UNDERNOURISHMENT ESTIMATES FROM THE TWO APPROACHES AND THE SOURCES OF THE DIFFERENCES **16**

a. Differences between the Underlying Parameters *17*

b. The Flaw in the Distribution based on CV_X Estimated on NHS household level data *19*

VI. CONCLUDING REMARKS **21**

REFERENCES **22**

SUMMARY

A non-parametric approach suggested by researchers from the International Food Policy Research Institute (IFPRI) for measuring food deprivation (undernourishment) is not an improvement to the current FAO parametric approach. This is mainly due to flaws arising from the use of an inappropriate methodological framework and the reliance on single household data from national household surveys (NHS) that are subject to undesirable sources of variation. FAO's parametric approach is still the only choice for estimating the prevalence of undernourishment for the purpose of monitoring hunger reduction at country, regional and global levels. The FAO approach estimates the average food consumption parameter from national food balances such as those from the FBS compiled and prepared by FAO on yearly basis. The FBS is the only data source for global monitoring. The parameter on inequality in food access is derived from NHS data, which are collected less frequently. For estimating the prevalence of undernourishment at sub-national levels and identifying population groups at high risk of food insecurity, countries are applying the FAO method to derive both the average and the inequality parameters from the NHS data.

I. INTRODUCTION

FAO has been traditionally estimating the prevalence of undernourishment in the total population using a parametric approach in the sense that it is based on the parameters of the distribution of dietary energy consumption (DEC) and a cut-off point reflecting an acceptable normative lower limit for dietary energy requirement (DER). This lower limit is referred to as the minimum dietary energy requirement (MDER). The part of the distribution of DEC below the MDER is taken as the estimate of the proportion of the population undernourished.

Recently, researchers from the International Food Policy and Research Institute (IFPRI) have proposed a non-parametric approach as an alternative to the FAO approach. The new approach is non-parametric in the sense that it is based on the direct comparison of the DEC of each sampled household in a NHS with the summation of the DER of all members in the corresponding household. The DER applied to each member is based on the median body-weight for the corresponding sex and age population group. Each household whose total DEC is below the respective total DER is classified as undernourished. The total number of individuals in the thus classified households is then divided by the total number of individuals in all the sampled households to estimate of the proportion of the population undernourished. This approach, which has been illustrated using NHS data for a number of countries in Sub-Saharan Africa, is being proposed by the IFPRI researchers as an improvement to the FAO approach (Smith, Alderman and Aduayom, 2006).

However the IFPRI researchers' proposal is misleading because of three main reasons. First, as the prevalence of undernourishment in the population is derived by comparing the DEC of each household in the sample with the DER obtained as an aggregation of the DER calculated for each of the individuals in the household, the resulting estimate is affected by the biases and errors inherent to the individual household level data from the NHS. A particular consequence of this approach is that it is implicitly based on biased estimates of the second moment of the distribution of DEC. This kind of bias is linked to the sampling designs used in NHSs (Scott, 1992; Arbia, 2002; Srivastava *et al*, 2002). Second, the calculation of the DER corresponding to each of the individuals in the household does not comply with the nutritional expert groups' recommendation that the energy requirements should be applied to groups and not single individuals of given sex and age (WHO, 1985; FAO, 2004). Third, the estimation of DER is incorrectly based on the 50^{th} percentile (median) of the distribution of acceptable body-weights for a given sex and age group. The use of the 50^{th} percentile leads to high probability of misclassifying normal individuals as undernourished.

The biases and errors that the household level data from the NHS are subject to leads to an overestimated inequality in DEC while the use of the 50^{th} percentile of the distribution of acceptable body-weights leads to overestimated DER values. As the effect of both is to raise the prevalence of undernourishment, it follows that the proposed non-parametric approach actually leads to overestimates. For this reason the estimates resulting from application of the approach gives the wrong impression that FAO's approach underestimates the prevalence of undernourishment.

This paper discusses the two approaches in the light of the above issues. Thus, the procedures involved in the two approaches are described in Sections II and III respectively. In Section IV, the inappropriateness of the non-parametric approach is discussed. Section V discusses the sources of the difference between the estimates of the prevalence of undernourishment resulting from the application of the two approaches for 12 countries in Sub-Saharan Africa and illustrates the flaw in the distribution underlying the estimates obtained using the non-parametric approach and the likely overestimation of the prevalence of undernourishment that this entails. Finally, Section VI emphasizes that there is still no alternative to the FAO approach for estimating the prevalence of undernourishment in a population and therefore efforts towards improvement should concentrate on improving the estimates of the parameters needed for applying this approach.

It is hoped that the views expressed in this paper will be helpful to the community of researchers and practitioners involved in food security assessments at the global as well as national level in clarifying the methodological issues addressed by the FAO approach and hence avoid the use of the non-parametric approach as proposed by the IFPRI researchers for the purpose of measuring undernourishment.

II. THE FAO PARAMETRIC APPROACH

According to the FAO approach, the estimate of the prevalence of undernourishment in the population is formulated as follows:

$$\mathrm{p}U = \int_{x<r_L} f_X(x)\,dx \qquad \text{……………………………………………(1)}$$

where X is a random variable representing dietary energy consumption (DEC), $f_X(x)$ is the density function of X, R represents DER and r_L is an acceptable lower limit of the distribution of R, i.e. MDER.

a. Derivation of the Formula for pU

The formula given by (1) was originally derived by considering the probability distributions of DEC and DER, i.e. X and R (Sukhatme, 1961). The formulation of the estimate of the prevalence of undernourishment within a distributional framework is based on two considerations: the first is that the food consumption data from household surveys refer to a probability sample rather the totality of households in the population and the second is that the DER of an individual is unknown but is normatively specified as the average for population groups of given age and sex.

The fact that the food consumption data from the NHS refers to probability sample of households from the population and DER is specified as an average implies that the inference regarding the prevalence of undernourishment has to be considered at the population level within a probability distribution framework. The unit of the distribution is the average individual implied by the expression of population data on per person basis. In other words the distribution refers to units that are free of the effect of differences due to sex and age.

There are in fact three probabilities regarding the status of an observed value of *X* vis-à-vis the individual's value of *R*: the probability of the observed value being below the individual's value of *R,* i.e. *P(x<r)*; the probability of the observed value being in balance with the individual's value of *R,* i.e. *P(x=r)*; and the probability of the observed value being above the individual's value of *R*, i.e. *P(x>r)*. At the population level these probabilities are conceived as an average or expected value over the distribution of *X* as follows:

$$P(X<R) = \int_{-\infty}^{\infty} P(x<r)\, f_X(x)\, dx \ldots\ldots (2)$$

$$P(X=R) = \int_{-\infty}^{\infty} P(x=r)\, f_X(x)\, dx \ldots\ldots (3)$$

$$P(X>R) = \int_{-\infty}^{\infty} P(x>r)\, f_X(x)\, dx \ldots\ldots (4)$$

As explained in detail in a separate paper (Naiken, 2007), the above population level probabilities depends on whether the variation of *R* is random or systematic. If the variation is random, e.g. due to measurement or estimation error, the three probabilities reduce to the following:

$$P(X<R) = \int_{-\infty}^{\mu_R} f_X(x)\, dx \ldots\ldots (5)$$

$$P(X=R) = 0 \ldots\ldots (6)$$

$$P(X>R) = \int_{\mu_R}^{\infty} f_X(x)\, dx \ldots\ldots (7)$$

where μ_R is the average or mean of *R*.

The above means that the use of the mean of *R* as cut-off point in estimating the prevalence of undernourishment (i.e. *P(X<R)*) implies that the variation of *R* is considered to be random.

However, since the variance of R considered here refers to the true variation arising from systematic sources, namely bodyweight and physical activity, the implied distribution of R in fact represents the distribution of X in a population where everyone is in the state of energy adequacy or balance. This means that the distribution of R reflects the realization of the joint distribution of X and R, i.e.

$$f_R(r)=f_{XR}(x,r)$$

where $f_R(r)$ represents the density function of R and $f_{XR}(x,r)$ the joint density function of X and R.

The above implies that $P(x=r)=1$ for all x overlapping the range of R. Thus, since by definition $P(x<r)=1$ for all x below the the lower limit of the range of R and $P(x>r)=1$ for all x above the upper limit, the three population level probabilities are given as follows:

$$P(X<R) = \int_{-\infty}^{r_L} f_X(x)\, dx \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(8)$$

$$P(X=R) = \int_{r_L}^{r_U} f_X(x)\, dx \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(9)$$

$$P(X>R) = \int_{r_U}^{\infty} f_X(x)\, dx \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(10)$$

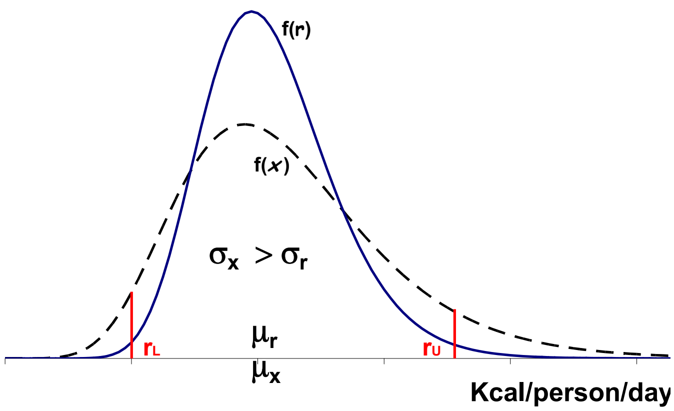
where r_L and r_U represent the lower and upper limits of the range of R.

Note that the right hand side (RHS) of (8) is equivalent to the RHS of (1). Thus the formula for pU given by (1) results from the consideration that the variation of R due to body-weight and physical activity is not random but systematic and consequently the implied distribution of R reflects the realisation of the joint distribution of X and R.

The probability framework illustrating the probabilities given by (8), (9) and (10) is shown in Figure 1.

In the figure, the distribution of X is shown to be wider than that of R since the distribution of requirement is located within the range of variation of X and the variance of X is expected to be larger than that of R. The larger variance of X is due the fact that it includes, in addition to the variance of R due to body-weight and physical activity, the variance due to income and residual factors. The area corresponding to $P(X=R)$, is represented by part of the distribution of X ranging from r_L to r_U while $P(X<R)$ is represented by the part below r_L and $P(X>R)$ by the part above r_U.

Figure 1: Distribution of R and X, $f_R(r)$ and $f_X(x)$, respectively



The means of X and R are shown to be equal in the figure solely for simplicity and the purpose of illustrating the extension of the distribution of X beyond the limits of the distribution of R. This implies a higher variance or standard deviation of X. It is obvious that in most cases, the two means are not equal.

As the extension of the two tails of the distribution of X beyond the limits of the distribution of R mainly reflects the effect of the income factor, the distribution of X is shown to be skewed to the right just as the income distribution.

The distribution of R also is likely to be skewed as it is induced by a slightly skewed distribution of weight for height in the reference population and the skewed distribution of physical activity levels which concentrates more population on the side of sedentary lifestyles than vigorous lifestyles. Moreover, as the true lower and upper limits of the range of R, i.e. r_L and r_U, are actually not known, the positions that they are shown in the figure reflect the fact that they have been taken to correspond to the 5^{th} and 95^{th} percentiles respectively of the distribution of R. In other words the 5^{th} and 95^{th} percentiles have been considered as acceptable limits of the range of R due to differences in body-weight and physical activity.

b. *Evaluation of the Formula for pU*

For the purpose of evaluating the formula for $P(X<R)$ it is necessary to specify the distribution of X, i.e. $f_X(x)$, and the lower limit of the distribution of R, i.e. r_L. In this context the distribution of X is assumed to be lognormal with parameters μ and σ. Thus