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

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:

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

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 capita calorie consumption outside of the range [Q1 – 2 x IQR, Q3 + 2 x IQR] are identified and consequently not used in the calculation of distributional parameters.

Although the IQR is a robust statistic and can also identify asymmetry in the data, the endpoints extend with equal distance on each side from the median, and for this reason, it is important to first symmetrize the data. Several different transformations exist for symmetrizing skewed data, and the most widely used of these is the log-transformation. Here the criterion used is the "medcouple", a robust measure of skewness insensitive to the presence of outliers and defined as a scaled median difference of the left and right sides of the

[4] This is especially the case for the SK parameter, which can be quite sensitive to the presence of extreme values [8].

distribution [10]. The medcouple was employed as a criterion for whether or not to apply the log transformation to improve symmetry. This was done by calculating the medcouple on the original and log-transformed data and comparing the two.

Two depictures of applications of Method 1, one in which the data was first symmetrized according to the medcouple rule and one in which the data was not, can be seen in Figures 4 and 5, respectively. The data before treatment can be seen in subfigures (a) for both, and the data after treatment is shown in subfigures (b). Figure 4 corresponds to data calculated from a 2005-06 NHS for Uganda while Figure 5 corresponds to a 1997-98 NHS for Vietnam. In the atypical case of Vietnam, the data was not log-transformed before application of IQR data treatment method as the log-transformation actually worsens symmetry of the data. In both cases, it is shown how data on both tails of the distribution has been treated. The percentage of outliers removed for these two countries is 0.8% for the Uganda dataset and 1.6% for the Vietnam dataset.

Another method used to assess the robustness of statistics for a sample uses what is known as the leave-one-out cross-validation approach (referred to as **Method 2**) [11]. With this approach, for a sample of size *n*, jackknife samples of size *(n – 1)* are created in which each observation is systematically left out of one subsample. For each subsample, the sensitivity of the statistic of interest to the excluded observation can be analyzed. As the sample skewness is a distributional measure extremely sensitive to outliers, here we focus on the robustness of the skewness measure to individual observations.

After calculating the skewness of the log-transformed data for each jackknife subsample, we can see how sensitive the statistic is to each observation. Observations that have a large impact on the skewness of the original log-transformed sample are removed. This is accomplished by using the IQR of our set of skewness values calculated from each jackknife subsample and an appropriate distance parameter. Individual observations which cause the original sample skewness to fall outside of the range [Q1 – 5 x IQR, Q3 + 5 x IQR], where Q1, Q3, and IQR are the 25th percentile, 75th percentile, and interquartile range of the set of jackknife subsample skewness values, respectively, are consequently excluded for the calculation of distributional parameters.

Graphical depictions of the application of the leave-out-one method can be seen in Figures 6 – 9. In Figures 6 and 8, data before and after treatment is shown for country datasets derived from a 2005-06 Uganda NHS and a 2006 Myanmar NHS, respectively. Figures 7 and 9 are graphical representations of the skewness of the jackknife subsamples calculated for the two data sets. The horizontal axes correspond to the observation number and the vertical axes are for the skewness of each jackknife subsample with the corresponding observation excluded. Note that the observations have been ordered according to increasing per capita calorie consumption. The horizontal dotted lines in the figures are the acceptable range around the median.

The third and final method for data treatment that will be examined here (referred to as **Method 3**) involves the use of influence measures with a linear regression relating food consumption to income. The regression, along with the effect of the month in which the survey took place to allow for a seasonality adjustment can be written as:

$$\boldsymbol{PPC_i = \beta_0 + \beta_1 * \log(inc_i) + \beta_2 Month_{1,i} + \beta_3 Month_{2,i} + \cdots + \beta_m Month_{m-1,i}} \quad (4)$$

where PPC_i is the per capita calorie consumption for household *i*, β_0 is an intercept term, β_1 is a regression parameter defining the linear relationship between the log of income and food consumption, and $Month_{j,i}$ is an indicator variable with value 1 if the survey for household *i* took place in month *j*.

A number of measures can be calculated after performing the regression to indicate how influential each observation is on the results. Using the R function *influence.measures()*, which is a part of the standard *stats* package, several different standard measures of influence can be calculated and whether or not an observation is influential can be determined [12]. The measures outputted from the function look at the effect each observation has on the regression coefficients and their precision, the fit of the linear regression, and the leverage of observations (both in terms of the value of the dependent and independent variables), and the decision as to whether each observation is influential is made based on standardized criteria [13]. Using this approach, those observations which are determined to be influential are excluded for the calculation of the SK and CV.

Examples demonstrating the application of the influences measures approach to data treatment can be seen in Figures 10, 11, and 12. In Figure 10, the relationship between the log of income and per capita calorie consumption is shown for an example with information from a 2005 Nicaragua NHS with the influential observations colored in green. Figures 10(a) and (b) are the same except that the range for the x-axis has been changed in (b). Figure 11 is intended to show the different parts of the overall process for calculation of the distributional parameters from the household survey data. Figure 12(a) shows the original data, (b) shows the data after the influential measures treatment (from which the SK is calculated), (c) shows the effect of using the income-consumption relationship, and (d) shows the seasonality adjustment. Figure 12 follows a similar logic as in Figure 11, except that the effect of the influence measures treatment method is applied at the end to highlight the affect that the treatment has on the data already adjusted by the linear regression.

A comparison of the percentage of observations removed using each of the three methods for a sample of 48 country datasets already used in *The State of Food Insecurity in the World* can be seen in Table 1. The average percentage of observations removed for the three methods are 1.7% for Method 1, 1.1% for Method 2, and 10.3% for Method 3. Method 2 is the method with the smallest standard deviation of the percentage of observations removed (the standard deviations of the percentage of observations removed across datasets are 1.41 for Method 1, 0.48 for Method 2, and 1.30 for Method 3). This gives some indication that the 2nd method is the most conservative and is the most stable across country datasets, and for this reason, Method 2 has been chosen as the default method for data treatment. This method allows a robust calculation of the SK parameter that is insensitive to any single observation found in the dataset.

For each method, as well as for the untreated data, the resulting CV and SK values can be seen in Table 2. In the last row, the root mean square difference of the parameters is shown for each method as compared to the parameters from the untreated data. The data treatment methods result in substantial differences in the SK parameter, highlighting the sensitivity of this parameter to extreme values. The CV is relatively more stable across data treatment methods, demonstrating the importance of methods to control for excess variability using grouping variables (described later) for the calculation of this parameter.

In order to further assess the performance of each of the data treatment methods and to examine the different criteria for the choice of the functional form of the distribution from section II, we will cross-examine the two in unison. Table 3 compares the PoU that results from applying each of the distributional algorithms to the CV and SK from each data treatment method. The total standard deviation of the PoU across distributional algorithms while holding the data treatment method constant is 0.026; on the other hand, the total standard deviation across the data treatment methods while holding the distributional algorithm constant is 0.022, implying that the functional form for the distribution has more of an impact on the PoU than the data treatment method.

ii. Capturing Excess Variability

The inequality in food access parameters are calculated using National Household surveys, from which the distribution of per capita calorie consumption within a country is formed with the aid of food composition tables [14]. More specifically, the calculation of the CV in food consumption is broken down into two parts, using the CV due to income (CV│y) and the CV due to all other factors orthogonal to income (CV│r):

$$CV = \sqrt{(\mathrm{CV} \mid \mathrm{y})^2 + (\mathrm{CV} \mid \mathrm{r})^2} \tag{5}$$

This procedure is applied in an attempt to capture habitual consumption from food acquisition data and as another level to control for noisy data.

The CV│y is calculated directly from survey data, when available, using a relationship between income and food consumption to control for excess variability. A previous comparison of the CV│r over time and across countries lead to setting the parameter equal to the constant value of 0.2. The parameter has been re-calculated here using demographic data on the sex and age class structure of the country. For those countries for which survey data have not been made available, indirect measures, using relationships between the CV│y and macroeconomic variables, are used to provide estimates [15].

As the original purpose of NHS is to measure the levels and changes in living conditions of the population, the data collected typically pertain to food acquisition over a given reference period. However, the aim of the food security analyses in this report is to capture habitual food consumption, which is expected to be less variable than food acquisition. Therefore, excess variability is controlled by assuming a stable relationship between income and consumption in calories, which nets out excess variability caused by some households boosting their food stocks while other households deplete theirs. In the past, this control for excess variability has been accomplished by grouping household food consumption according to income deciles to calculate the CV│y:

$$CV_{INC} = \sqrt{\frac{w_j \sum_{j=1}^{10} [\mu_j - \mu]^2}{\sum_{j=1}^{10} w_j}} \tag{6}$$

where μ is the population weighted mean in per capita calorie consumption, μ_j is the population weighted mean in calorie consumption of the j^{th} income decile, and w_j is the sum of the population weights of all observations in the j^{th} income decile. This method has been used for quite some time to control for excess variability, and a more in depth discussion can be found in [15].

As an extension of the method described above, a linear regression as equation 4, linking the log of per capita income to per capita calorie consumption, can also be used, along with indicator variables for the month in which the survey was conducted to control for seasonality. The CV│y is then calculated from the fitted values from the regression adjusted for seasonality. This approach for controlling for excess variability is similar to the one using income deciles, but the main advantage of the regression approach is that an adjustment can be made for seasonality, for the months in which the survey was conducted.

Table 4 shows the relationship among the different methods for controlling for excess variability. As can be seen in the table, the correlation between the empirical CV and the coefficient of variations derived from each of the different control methods is relatively low, whereas the relationship between the CV│y calculated by grouping by income decile and those calculated using the predicted values from a regression with log of income is quite high. As the calorie and log of income regression is an improvement to the income decile method for controlling for excess variability that allows for some control of seasonality, it has been chosen.

For datasets analyzed in partnership with National Statistical Offices, a high correlation was found for the derived CV values between the data treated according to the country and FAO methods. The correlation between the derived CV│y values calculated using data treatment Method 2 from section V and country methods was 0.86 for the when using income decile to control for excess variability and 0.83 when using the regression with log of income. On the other hand, the empirical skewness values (corresponding to the skewness parameter for the PoU) have a correlation of only 0.002 between the data treatment Method 2 and country methods. For this reason, the skewness values from surveys for which it is impossible to apply data treatment Method 2 (because of the unavailability of untreated data) have not been used, and the SK parameter has been derived according to relationship (1).

iii. Update of indirect Coefficients of Variation, for countries where surveys are not available or reliable

The procedure described so far is employed in countries where one or more reliable NHS are available. Where this is not the case, so-called indirect estimates for the variability in food consumption are used. Indirect estimates for the variability in food consumption due to income were last updated for countries for which no survey data is available by using relationships between estimates of the CV from available household survey data, the Gini coefficient of income (Gini), GDP, and infant mortality [15]. In the past, the PoU indicator methodology was frequently criticized for holding CVs – which account for inequality in food consumption – constant over time for most countries [16]. Since then, more household surveys have been made available from which the CV in food consumption has been estimated. Using these CV values, we can build more accurate macroeconomic relationships to provide more precise updated indirect estimates. In addition, much more detailed macroeconomic information has been made available (especially for Gini coefficients), again increasing the precision of updated indirect estimates and allowing them to vary over time.

To update relationships used previously for indirect estimates with this new information, the most comprehensive dataset available of Gini coefficients [17] has been used in a regression to relate the variability in food consumption due to income to the Gini and the log of GDP. To ensure cross-country comparability in different points in time, per capita GDP in constant 2005 international dollars in purchasing power parity terms, calculated by the World Bank, has been used. Regional indicators have been included for Africa, the Americas, Asia, and Western Asia. These regional indicators are at the macro level, except that Western Asia was separated from the rest of the continent. This decision was made based on data availability to allow for an adequate number of data points within each category.

Since there are multiple observations (i.e. more than one survey) for some countries, a weighted regression was used in which each observation is weighted by the one over the number of surveys for that country. The variables in the regression have been scaled by their standard deviations to facilitate interpretation of the output, and the regional dummy variables are presented relative to Africa. The results in Table 5 show the updated version of the relationship previously estimated, and it is shown that the Gini and log of GDP are both significant in explaining the variability in the CV│y. The magnitudes of the parameter estimates give an idea of the relative importance of the variables on inequality in food consumption. As can be seen in the results, a higher Gini is associated with a higher inequality in food consumption and a higher GDP is associated with a lower degree of inequality in food consumption. We will now include another important aspect of inequality in food consumption into the regression, namely the effect of the relative price of food.

To investigate in full the effects of changes in food prices on inequality in food access, it is ideal to have a measure of local prices. In collaboration with the World Bank, FAO has developed a relative price of food indicator using data from the International Comparison Program [18] and consumer and food price indices available on FAOSTAT [19]. The indicator is designed to capture changes in domestic food prices that are comparable over time and among countries. The ratio of food and general consumption in purchasing power parity (PPP) terms is projected forwards and backwards in time using the ratio of the country's consumer food price index to the country's general consumer price index, relative to that of the United States of America.

The indicator can be written out as:

$$\boldsymbol{FPLI_{x,\ i} = \frac{Food\ PPP_{x,\ 2005}}{General\ PPP_{x,\ 2005}} * \frac{\frac{FPI_{x,\ i}}{FPI_{USA,\ i}}}{\frac{CPI_{x,\ i}}{CPI_{USA,\ i}}}} \quad \textbf{(7)}$$

where:

$Food\ PPP_{x,\ 2005}$ = Food and non-alcoholic beverages consumption in purchasing power parity terms in 2005

$General\ PPP_{x,\ 2005}$ = Actual individual consumption in purchasing power parity terms for country x in 2005

$FPI_{x,\ i}$ = The food price index for country x in year i with 2005 base year

$FPI_{USA,\ i}$ = The food price index for the United States in year i with 2005 base year

$CPI_{x,\ i}$ = The consumer price index for country x in year i with 2005 base year

$CPI_{USA,\ i}$ = The consumer price index for the United States in year i with 2005 base year

The indicator is available as part of a suite of food security indicators produced by the food security and social statistics team at the FAO and is available online [20]. The regional and country aggregates are built using the Purchasing Power Parity data for Gross Domestic Product in Current International Dollars from the World Bank as a weighting variable. A plot of the indicator over time for by region can be seen in Figure 13. The plot illustrates that the relative price of food for Africa and Asia is on the rise and much more volatile, while it is lower and more stable for other regions of the world.

Results from a regression incorporating the Relative Price of Food Indicator can be seen in Table 6. The GDP and relative price of food indicators are included on the log scale, implying that changes in these variables at low values will have a larger impact on the CV due to income. An interaction term between the GDP and the relative food price indicator has been included to allow for differential effects of the price of food at different levels of GDP. Table 7 shows the pair-wise correlation of the independent variables included in the regression – although there is a relatively high correlation between log(GDP) and the log of the relative price of food, the correlation is not high enough to cause multi-collinearity problems with the estimation of the model parameters. In Appendix 1, the CV | y values used in the regression (from data treatment **Method 2)**, along with the values of the relevant independent variables can be found[5].

In order to interpret the effect of GDP and relative food price, both the individual effect parameters and interaction parameter need to be considered. In this way, the average effect of GDP is obtained by summing the individual effect of GDP (0.0663) and the interaction effect with price (-0.171) to obtain the overall effect of GDP (-0.1047) – this means that an increase in GDP is associated with higher equality (or lower inequality) in food consumption. Likewise, the average effect of price is obtained by summing the individual effect of price with the interaction effect with GDP to obtain the overall price effect (0.125). The price effect works in the opposite direction of GDP to cause an increase in inequality in food consumption with a relative price increase.

With the parameters from the regression described above, the variability in food consumption due to income has been updated for countries with available Gini coefficients and available data on the relative price of food and GDP. Note that the Gini coefficients in the World Bank database differ in terms of whether they are calculated with reference to the household or the individual, consumption or expenditure, and gross or net income – these differences can make comparability across different types of Gini coefficients difficult [21]. For this reason, care was taken to ensure that the same type of Gini calculation was used within a single country and, to maintain cross-country comparability, only relative changes in the predicted values from the regression were used to update the CV parameter. The resulting updates take into account economic progress in a country as well as changes in relative food prices, allowing for a more complete picture of inequality in food consumption.

iv. Variability Due to Factors different from income

Recall that to obtain the total variability in food consumption used to calculate the PoU, the variability that is due to income ($CV \mid y$) is added to the variability due to all other factors that are not correlated with income($CV \mid r$):

$$CV(x) = \sqrt{(CV|y)^2 + (CV|r)^2}$$

Much of the variability orthogonal to income is due to differences in energy requirement, which are in turn largely determined by population structure as well as by physical activity levels, life styles, access to safe drinking-water, and progress in health care and disease reduction. Previous analyses showed small variability in this subcomponent across countries and over time, compared with the income component, and the variability due to requirement has been maintained at a fixed value [15]. Since this decision was made, the distribution of the world's population across sex and age class has become more uniform [22], and there is a need to take into account these changes.

To take into account the world's rapidly changing population structure, time-varying country estimates for the variability in food consumption due to requirement have been calculated. Using estimates for the average dietary energy requirement by sex and age class [7] and corresponding population ratios [23] as weights, the variance due to requirement is estimated for a given country in a given year. Further work is under way to capture the rest of the variability that is orthogonal to income. The revision discussed here allows estimates of the variability in food consumption to reflect more accurately demographic differences across countries and demographic evolution within a country.

[5] Note that Myanmar 2006 had to be left out of the regression, as information on the GDP of the country was not available from the World Bank.

VI. Limitations of the methodology and conclusion

The FAO methodology for estimating undernourishment has been subject to long-standing and wide debate. The methodology suffers from several limitations, which need to be acknowledged and taken into account when analysing the results presented in this report.

First, the indicator is based on a narrow definition of hunger, covering only chronically inadequate dietary energy intake lasting for over one year. Energy intake is a very specific aspect of food insecurity, which applies where conditions are more severe. Individuals experiencing difficulties in obtaining enough food are likely to switch towards cheaper sources of energy and to compromise the quality of their food intake in a way that can create substantial damage [24]. To address this limitation, the FAO suite of food security indicators has been presented since the 2012 edition of *The State of Food Insecurity in the World*. The suite comprises indicators that reflect a broader concept of food insecurity and hunger and allows consideration of their multifaceted nature.

Second, the PoU indicator cannot capture within-year fluctuations in the capacity to acquire enough energy from food, which may themselves be causes of significant stresses for the population. Within-year fluctuations can also affect the quality of the diet, as consumers will resort to cheaper foods during periods when access becomes more difficult.

Third, the FAO methodology for computing undernourishment cannot take into account any bias that may exist in intra-household distribution of foods [25], such as that arising from cultural habits or gender-based habits and beliefs. As seen, the parameters that describe the distribution of food across the population are derived from household-level surveys, rather than from information that refers to individuals.

A final and significant limitation of the FAO methodology for computing the prevalence of undernourishment is that it does not provide information on the degree of severity of the food insecurity conditions experienced by a population. The parametric model described in this annex allows only estimates of the undernourished share in a population, but is essentially silent about the composition of undernourishment within that part of the population.

Within the debate on measuring undernourishment, the FAO methodology has frequently attracted two critiques:

The indicator underestimates undernourishment, as it assumes a level of physical activity associated with a sedentary lifestyle, while poor people are often engaged in physically demanding activities.

The methodology is based on macrodata, whereas microdata from surveys allow accurate measurement of food consumption.

The first criticism is unfounded. Ideally, undernourishment should be assessed at the individual level by comparing individual energy requirements with individual energy intakes. This would enable the classification of each person in the population as undernourished or not. However, this approach is not feasible for two reasons: individual energy requirements are practically unobservable with standard data collection methods; and individual food consumption is currently measured with precision in only a few countries and for relatively limited samples. The individual-level consumption data that can be estimated from NHS are largely approximated owing to disparities in intra-household food allocation, the variability of individual energy requirements, and the day-to-day variability of food consumption that can arise for reasons independent of food insecurity. The solution adopted by FAO has been to estimate the PoU with reference to the population as a whole, summarized through a representative individual, and to combine available microdata on food consumption with macrodata. Within the population, there is a range of values for energy requirements that are compatible with healthy status, given that body weight, metabolic efficiency and physical activity levels vary. It follows that only values below the minimum of such a range can be associated with undernourishment, in a probabilistic sense. Hence, for the PoU to indicate that a randomly selected individual in a population is undernourished, the appropriate threshold is the lower end of the range of energy requirements.

As for the second criticism, the FAO methodology in fact combines available microdata on food consumption derived from surveys with macrodata from food balance sheets. Food balance sheets provide information on the amount of food that is available for consumption after taking into account all the possible alternative uses of the food items; hence, they provide approximate measures of per capita consumption, which are available for a large number of countries and are homogenous. The methodology adopted for computing these data is currently under revision, together with the estimates of waste parameters employed to derive the DEC, so the level of accuracy is expected to increase in the next few years. Survey data, where available and reliable, are employed in the FAO methodology to compute the variability (CV) and skewness (SK) parameters that characterize the distribution of food consumption $f(x)$. It is therefore essential that surveys are improved to obtain more accurate measures of undernourishment. Such improvement will require promoting both greater standardization across existing surveys, particularly household budget surveys, and conducting more refined surveys that capture food intake at the individual level. At the moment, few surveys accurately capture habitual food consumption at the individual level and collect sufficient information on the anthropometric characteristics and activity levels of each surveyed individual; in other words, very few surveys would allow estimation of the relevant energy requirement threshold at the individual level.

To conclude, the quality of the PoU estimates depends heavily on the quality of the background data employed in the estimation. Hence, to obtain better estimates on undernourishment it is important to improve food consumption data through the design and implementation of high-quality nationally representative surveys that are comparable over time and across countries. Improvements in currently available surveys are critical to obtain more accurate measures of undernourishment through the FAO methodology. This requires both promoting a higher level of standardization across existing surveys, and particularly household budget surveys, as well as the promotion of more refined surveys that can capture food intake at the individual level. At the moment, there are few surveys that capture accurately habitual food consumption at the individual level and collect sufficient information on the anthropometric characteristics and activity levels of each surveyed individual; in other words, very few surveys would allow estimating relevant energy requirement threshold at the individual level.

Despite these criticisms, the Prevalence of Undernourishment indicator is an invaluable tool to monitor progress towards reducing global hunger. As with any monitoring indicator, the harmonious and uniform application of a well-established methodology is crucial. Nonetheless, innovations to the methodology, such as those presented in previous versions of *The State of Food Insecurity in the World* , as well as the further refinements presented here can help us to more accurately capture how progress is being made in reducing hunger and how the problem is currently distributed globally. The first refinement presented in this paper has been a data-driven flexible selection criterion for the choice of the functional form of the distribution of per capita habitual calorie consumption that maintains the probability framework. Further improvements to the calculation of inequality in food access parameters, both directly and indirectly, have been made to allow for time-varying parameters that take into account economic progress and demographic changes. These refinements, taken together, provide more accurate estimates of the Prevalence of Undernourishment for better-informed policy.

Figure 1: PoU for Function 1

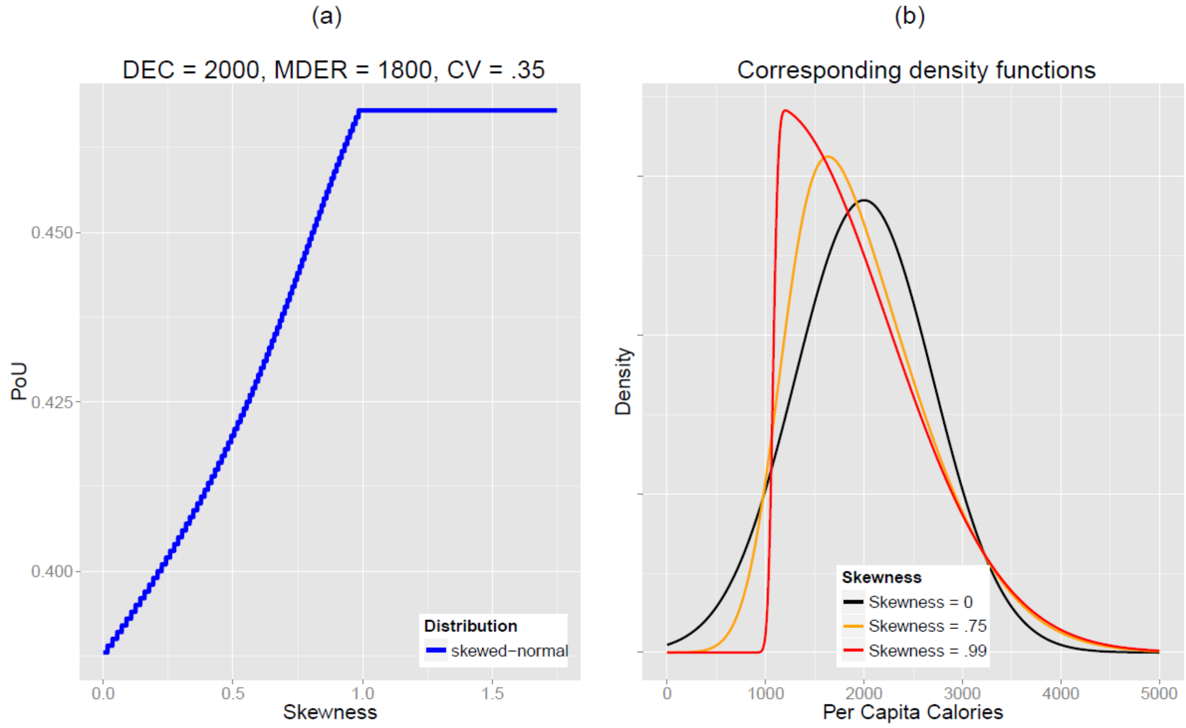


Figure 2: Evolution of the distribution for Function 3 with switch parameter .4

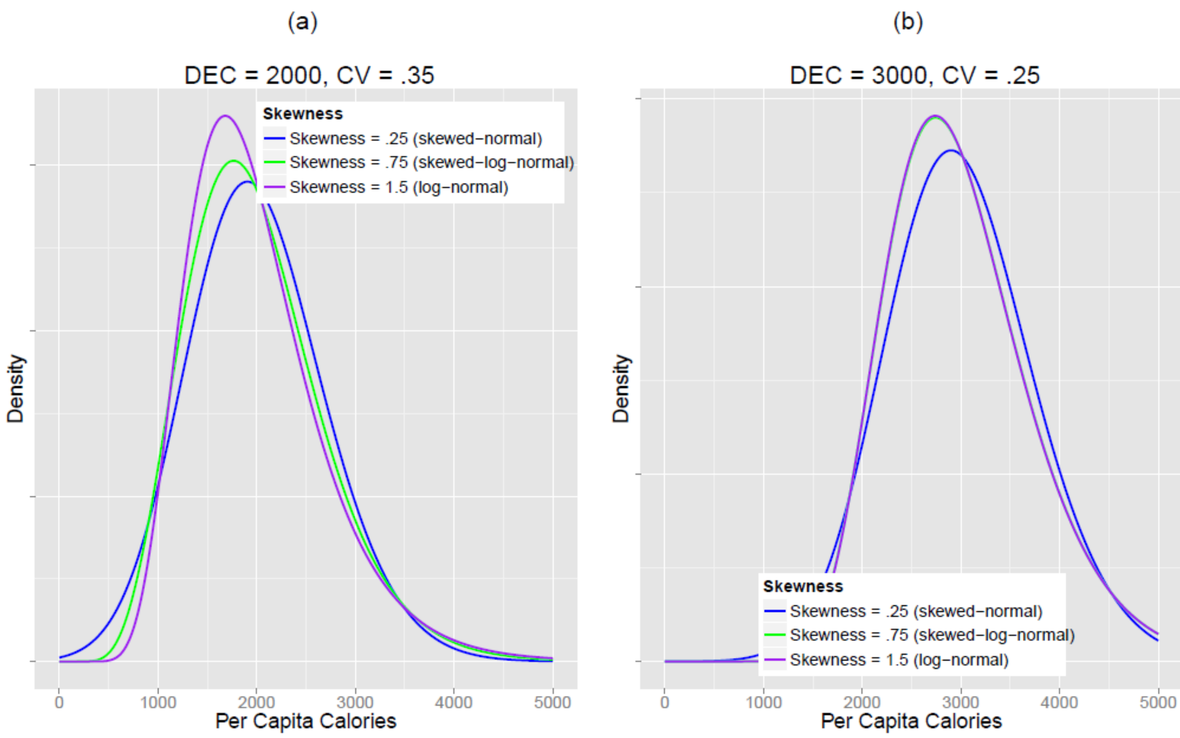
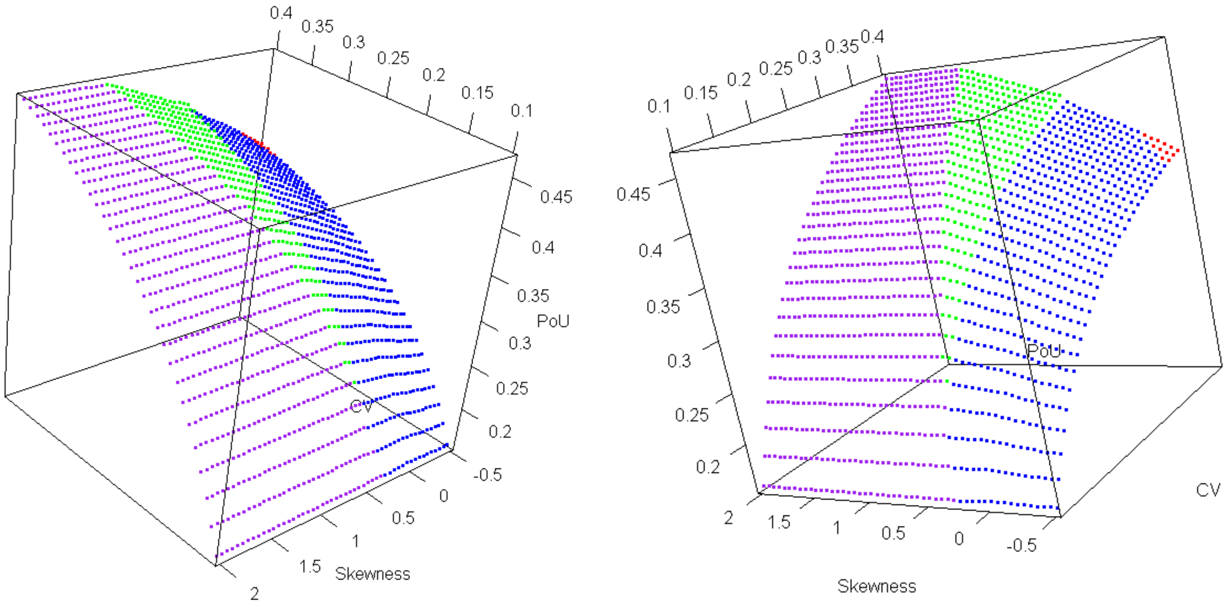


Figure 3: 3D plots for Function 3

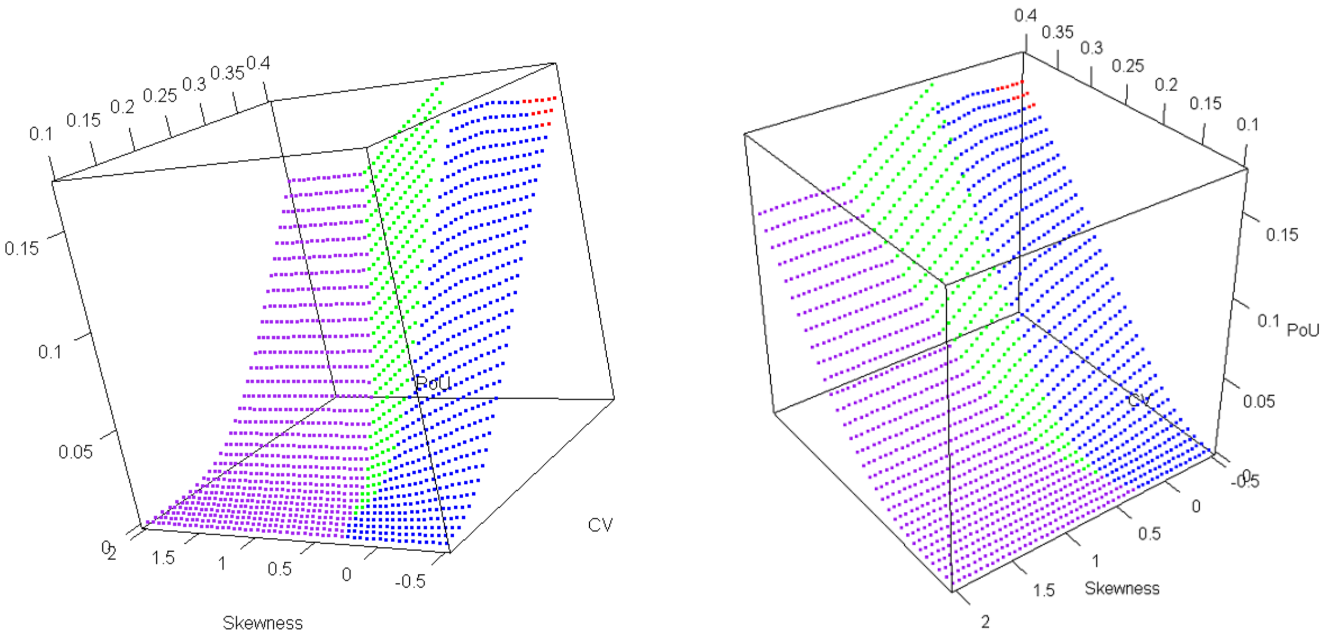
3D plot of PoU vs. CV and Skewness with DEC fixed at 2000, MDER at 1800

(a) **(b)**



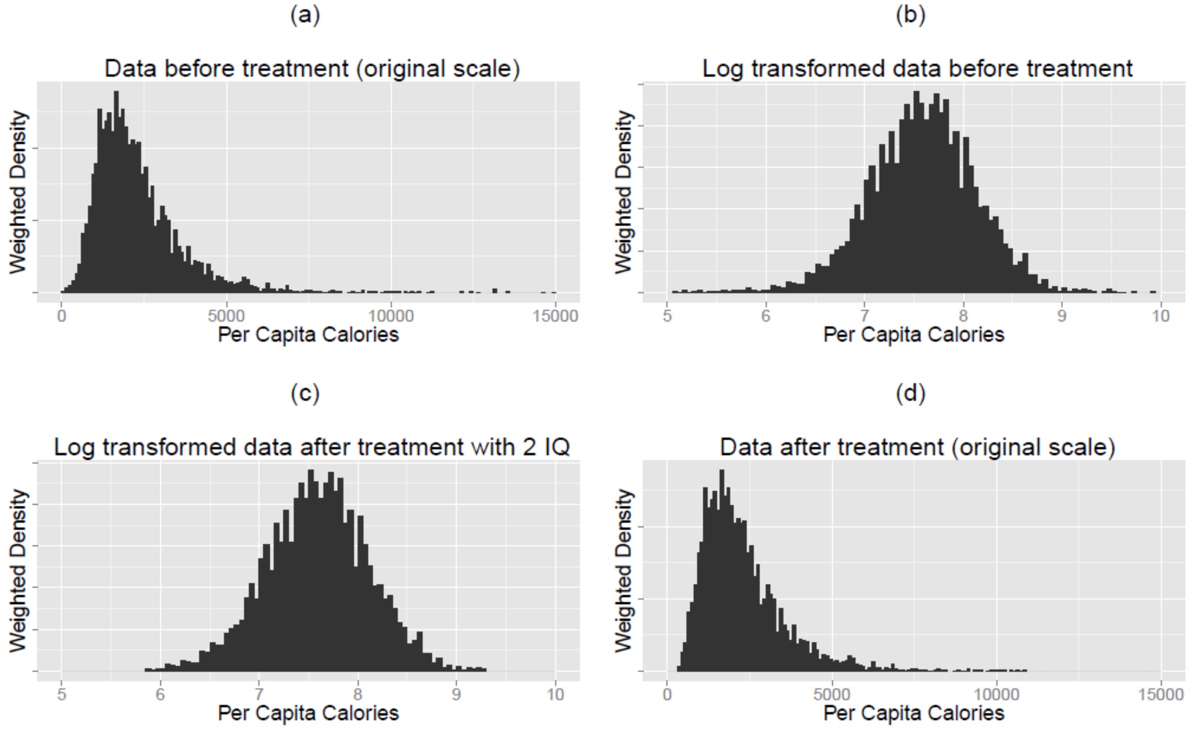
3D plot of PoU vs. CV and Skewness with DEC fixed at 3000, MDER at 1800

(c) **(d)**



In the red colored area, there is significant probability of negative per-capita calorie consumption (probability greater than .01). Parameters derived from HIES do not fall into this case.

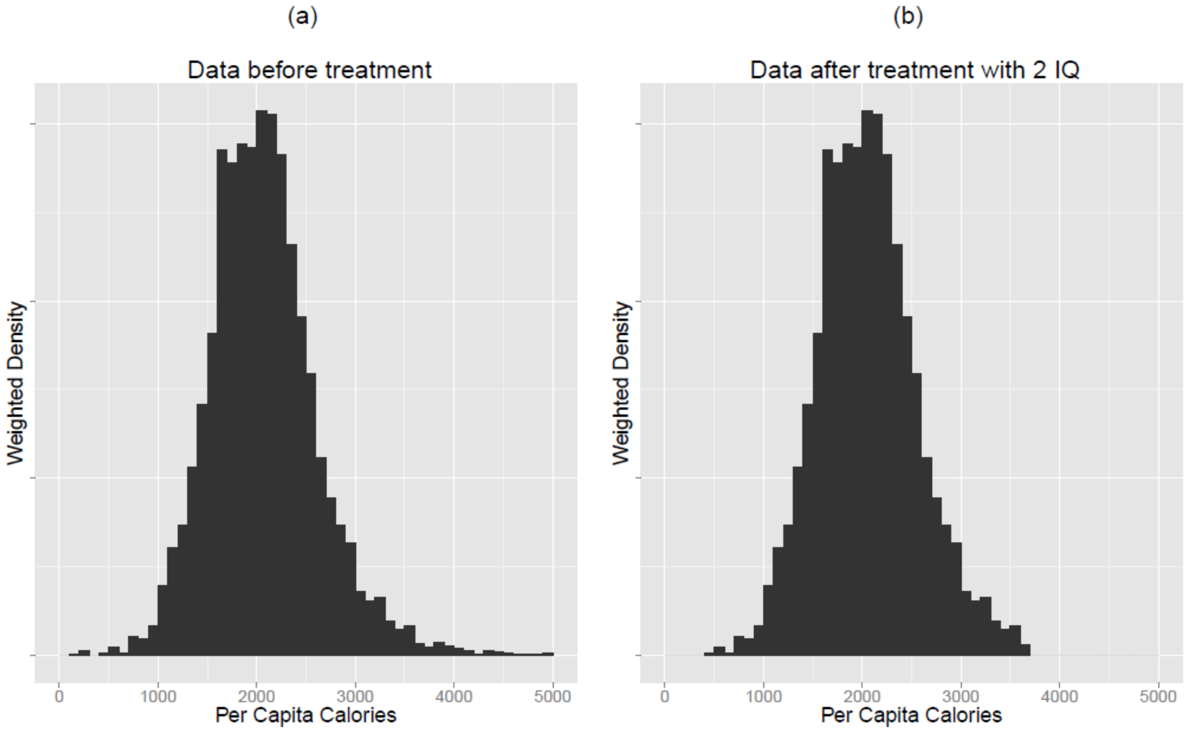
Figure 4: Data treatment with 2* IQ (Method 1) for Uganda 2005-06 with log transform



Empricial Weighted Statistics:

(a) Mean: 2265, CV: .58, SK: 2.78 (b) Mean: 7.58, CV: .071, SK: -.250
(c) Mean: 7.59, CV: .068, SK: -.076 (d) Mean: 2257, CV: .54, SK: 1.67

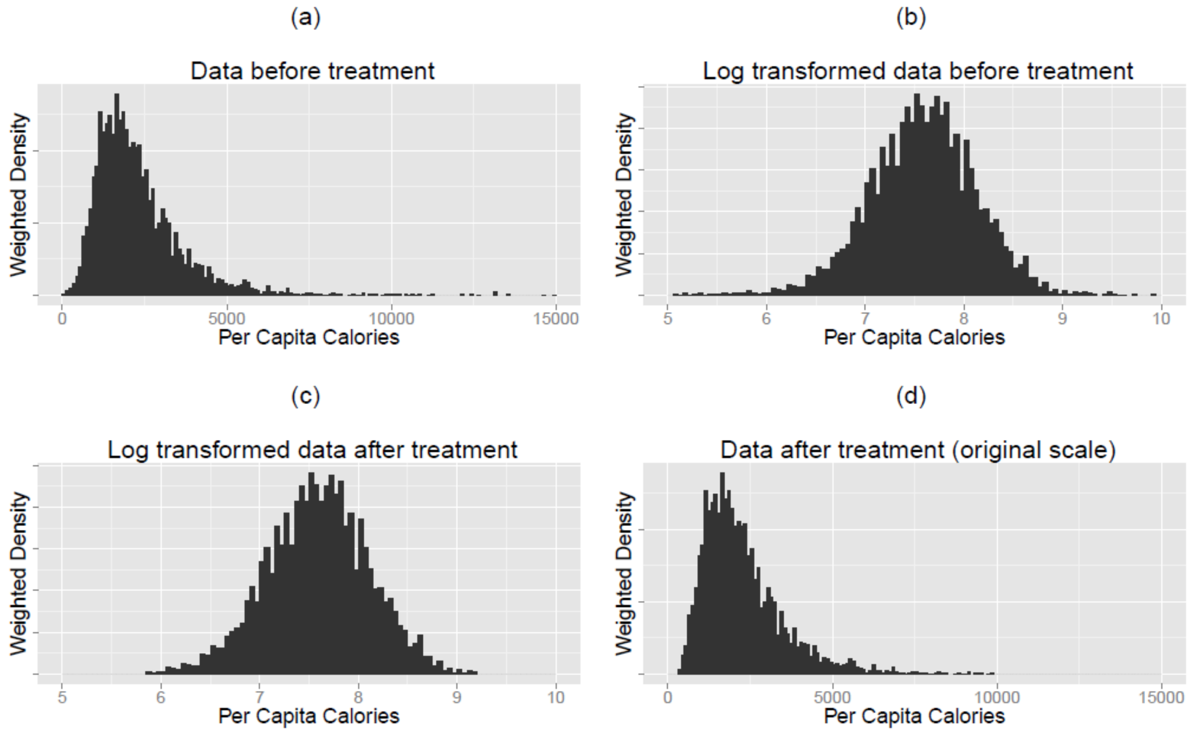
Figure 5: Data treatment with 2 * IQ (Method 1) for Vietnam 1997-98



Empricial Weighted Statistics:

(a) Mean: 2076, CV: .28, SK: 1.79 (b) Mean: 2052, CV: .25, SK: .30

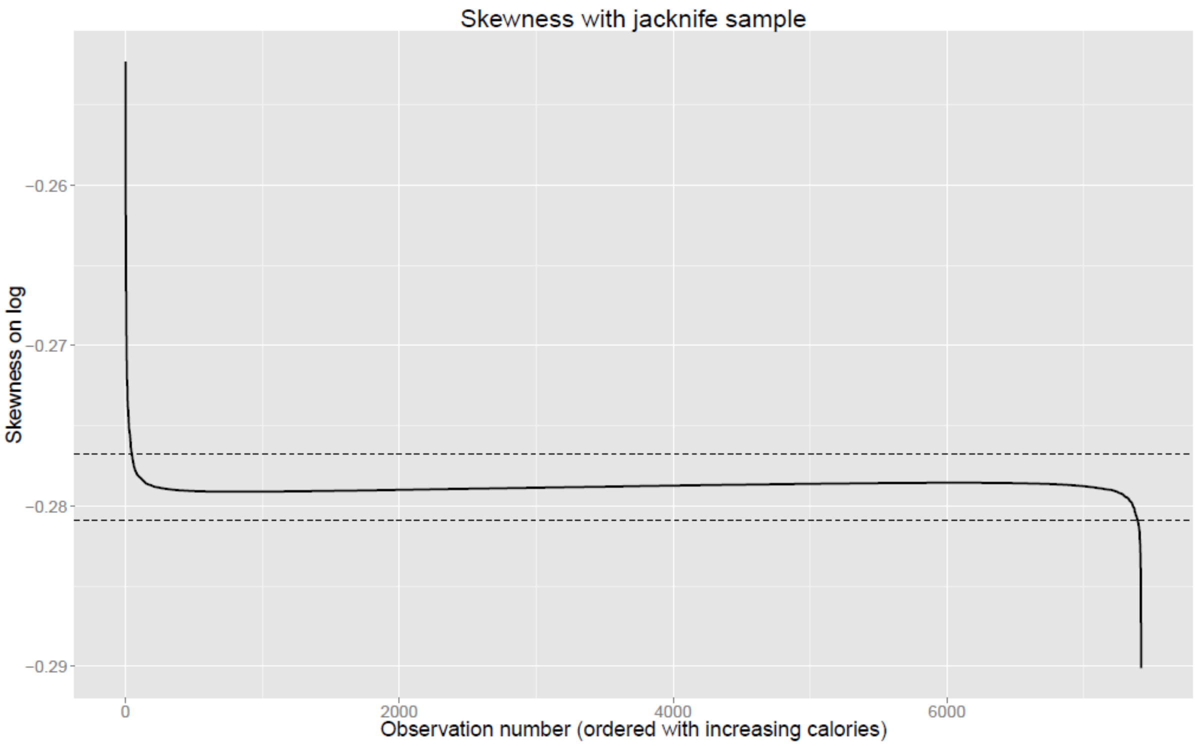
Figure 6: Data treatment using skewness sensitivity (Method 2) for Uganda 2005 06



Emprical Weighted Statistics:

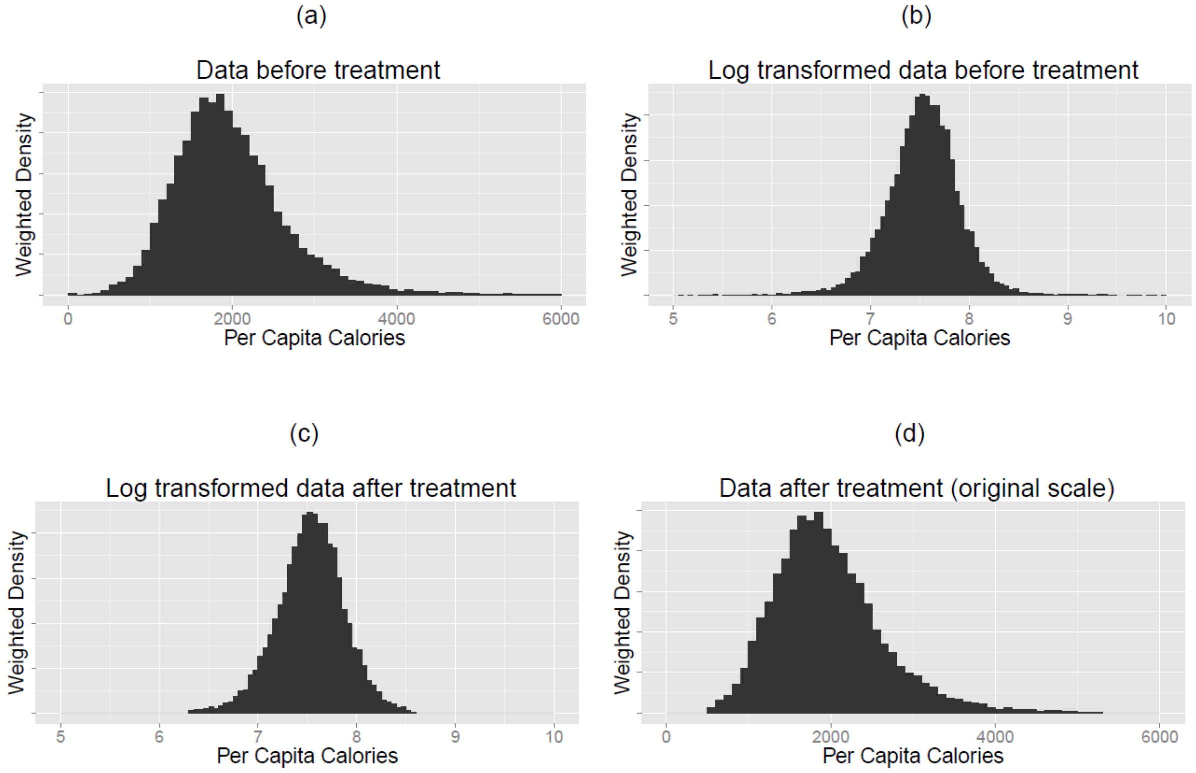
(a) Mean: 2265, CV: .58, SK: 2.78 (b) Mean: 7.58, CV: .071, SK: -.250
(c) Mean: 7.59, CV: .068, SK: -.099 (d) Mean: 2249, CV: .53, SK: 1.50

Figure 7: Plot of skewness sensitivity (Method 2) for individual observations for Uganda 2005-06



Note: Dotted lines depict 3 times the interquartile range around the median skewness from jacknife samples

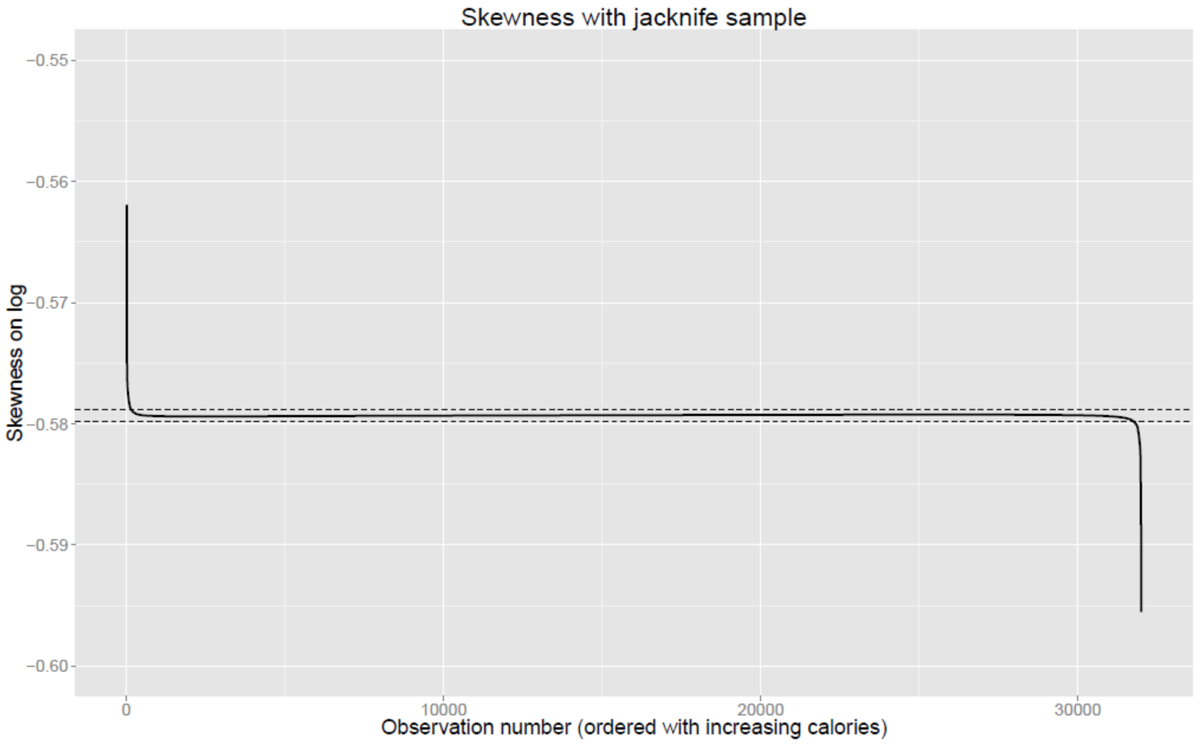
Figure 8: Data treatment using skewness sensitivity (Method 2) for Myanmar 2006



Empricial Weighted Statistics:

(a) Mean: 1995, CV: .47, SK: 17.8 (b) Mean: 7.53, CV: .051, SK: -1.47
(c) Mean: 7.53, CV: .045, SK: -.224 (d) Mean: 1971, CV: .34, SK: .94

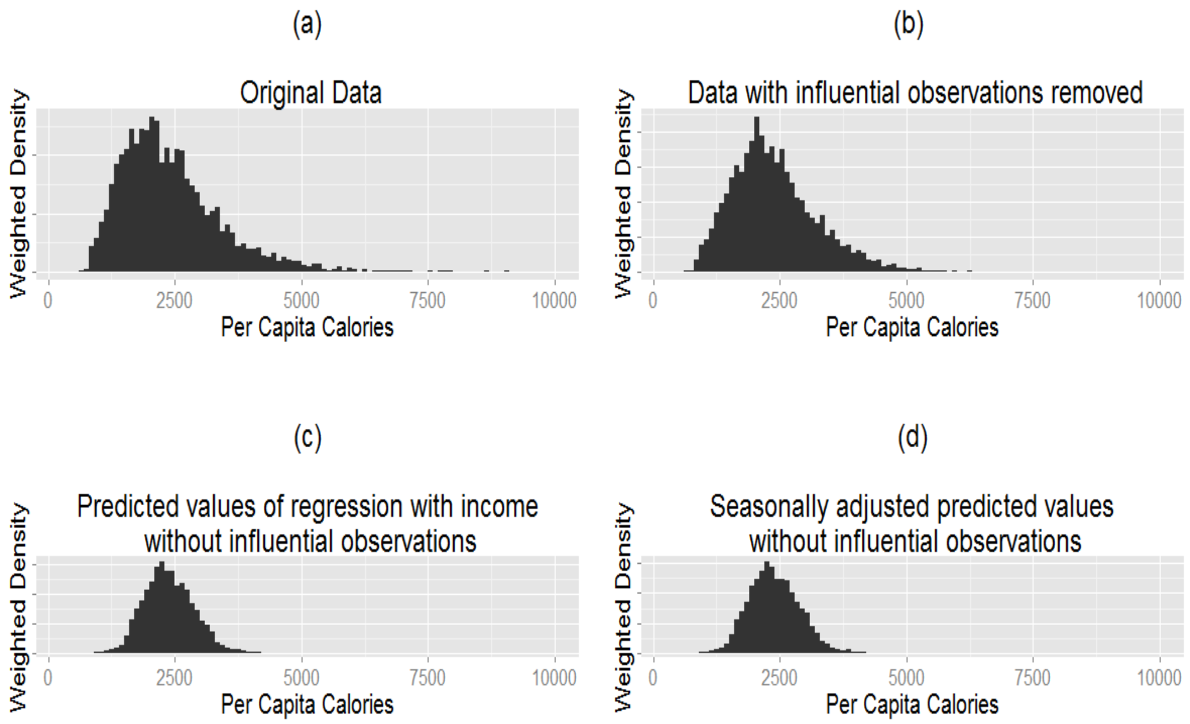
Figure 9: Plot of skewness sensitivity (Method 2) for individual observations for Myanmar 2006



Note: Dotted lines depict 3 times the interquartile range around the median skewness from jacknife samples

Figure 10: Influential observations (Method 3) for Nicaragua 2005

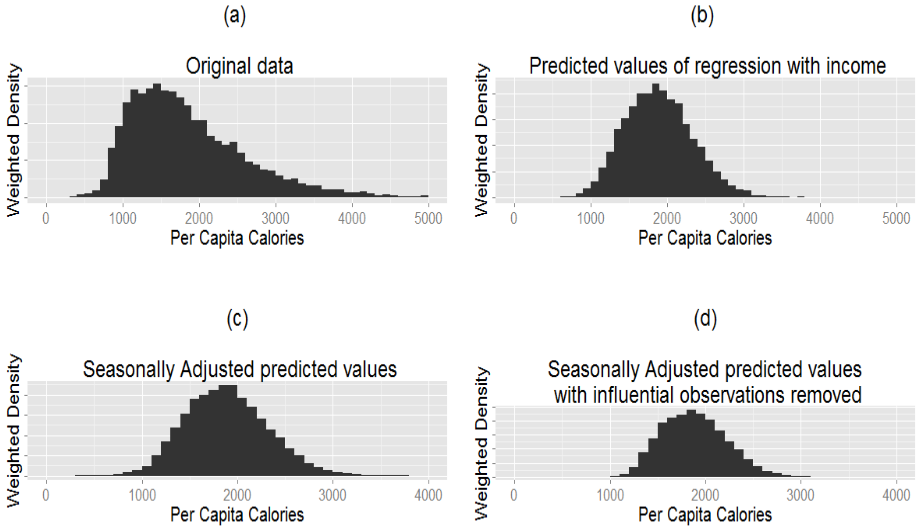
Figure 11: Influential data treatment process (Method 3) for Nicaragua 2005



Empricial Weighted Statistics:

(a) Mean: 2440, CV: .45, SK: 2.58
(b) Mean: 2393, CV: .36, SK: .93
(c) Mean: 2393, CV: .21, SK: .32
(d) Mean: 2397, CV: .21, SK: .30

Figure 12: Influential Data treatment process (Method 3) for Mongolia 2007-08



Empricial Weighted Statistics:

(a) Mean: 1890, CV: .63, SK: 26.04 (b) Mean: 1890, CV: .24, SK: .36
(c) Mean: 1890, CV: .23, SK: .36 (d) Mean: 1884, CV: .18, SK: .34

Table 1: Percentage of observations removed for the 3 data treatment methods

Country	Year	Method 1 (%)	Method 2 (%)	Method 3 (%)	Country	Year	Method 1 (%)	Method 2 (%)	Method 3 (%)
Albania	2005	0.3	0.6	11.0	Mexico	2004	1.8	1.2	10.1
Azerbaijan	2006	1.7	1.2	11.7	Mexico	2006	1.5	1.0	10.5
Bangladesh	2000	1.2	1.3	8.8	Mexico	2008	0.9	0.9	11.2
Bangladesh	2005	1.4	1.4	7.8	Moldova	2006	0.6	0.6	10.7
Bangladesh	2010	1.1	1.2	9.5	Mongolia	2007	0.4	0.6	8.1
Bulgaria	2001	1.1	1.1	9.2	Mozambique	2002	1.8	1.2	10.4
Burkina Faso	2003	1.1	1.1	9.1	Myanmar	2006	1.4	1.2	10.5
Cambodia	2003	4.7	2.5	8.9	Nepal	1995	1.0	1.1	8.9
Cambodia	2009	7.3	2.6	9.3	Nicaragua	2005	0.7	0.7	12.3
Chad	2009	1.3	1.2	8.4	Pakistan	2005	1.1	1.0	10.7
Congo, DR	2004	0.9	1.0	12.1	Panama	2008	3.4	1.3	11.1
Egypt	1997	2.2	2.0	8.8	Papua NG	1996	0.7	0.7	12.4
Ethiopia	1999	0.1	0.2	13.5	Paraguay	1997	2.2	1.4	10.9
Georgia	2005	1.4	1.0	11.0	Peru	2003	3.8	1.3	9.6
Guatemala	2006	1.0	1.0	11.5	Philippines	2003	0.6	0.8	9.8
Hungary	2002	0.7	0.6	11.7	Sri Lanka	1999	1.3	1.2	10.2
India	2000	1.2	1.1	9.2	Sudan	2009	3.7	0.9	11.0
India	2004	1.7	1.4	11.1	Tajikistan	2007	1.2	1.0	10.1
Indonesia	2008	0.2	0.4	9.9	Tanzania	2000	1.3	0.8	10.4
Iraq	2007	3.8	2.2	10.9	Tanzania	2007	5.6	0.9	13.2
Cote d'Ivoire	2002	1.7	1.3	10.2	Thailand	2011	1.5	1.2	11.6
Lao PDR	2002	0.8	0.2	8.2	Uganda	2005	0.8	1.0	10.0
Lao PDR	2007	3.2	1.8	8.7	Vietnam	1997	1.6	1.0	9.8
Lithuania	2002	0.4	0.5	11.0	Vietnam	2006	1.4	1.1	9.4

Table 2: Empirical CV and SK values from the 3 data treatment methods

		No Treatment		Method 1		Method 2		Method 3	
Country	Year	CV	SK	CV	SK	CV	SK	CV	SK
Albania	2005	0.57	2.53	0.55	1.60	0.54	1.44	0.46	0.94
Azerbaijan	2006	0.23	1.06	0.21	0.64	0.21	0.81	0.19	0.79
Bangladesh	2000	0.29	1.65	0.26	0.83	0.26	0.80	0.24	0.44
Bangladesh	2005	0.30	2.66	0.26	0.92	0.26	0.91	0.24	0.47
Bangladesh	2010	0.30	2.11	0.27	0.91	0.27	0.96	0.24	0.59
Bulgaria	2001	0.57	3.39	0.51	1.74	0.50	1.52	0.44	0.90
Burkina Faso	2003	0.96	17.56	0.66	2.21	0.67	2.40	0.58	1.88
Cambodia	2003	1.51	39.02	0.42	1.44	0.47	1.90	0.48	2.29
Cambodia	2009	4.75	53.64	0.53	1.88	0.75	3.55	1.04	6.25
Chad	2009	1.03	11.39	0.77	2.65	0.78	2.79	0.68	1.70
Congo, DR	2004	0.87	3.93	0.81	2.75	0.81	2.85	0.64	1.92
Egypt	1997	0.71	6.25	0.53	1.89	0.53	1.85	0.47	1.19
Ethiopia	1999	0.42	1.21	0.41	1.19	0.41	1.17	0.34	0.67
Georgia	2005	1.08	5.24	0.93	2.82	0.95	3.02	0.76	1.76
Guatemala	2006	0.58	2.61	0.54	1.68	0.54	1.54	0.47	1.13
Hungary	2002	0.54	2.93	0.50	1.64	0.50	1.79	0.44	1.10
India	2000	0.30	4.30	0.27	0.90	0.27	0.77	0.24	0.74
India	2004	0.41	21.51	0.26	0.86	0.26	0.75	0.25	0.82
Indonesia	2008	0.30	1.03	0.30	1.00	0.30	1.00	0.27	0.64
Iraq	2007	0.53	6.44	0.40	1.40	0.42	1.39	0.38	0.89
Cote d'Ivoire	2002	2.99	34.42	1.07	3.25	1.16	4.08	1.09	3.44
Lao PDR	2002	0.73	2.44	0.66	0.46	0.70	1.16	0.64	0.31
Lao PDR	2007	0.56	6.26	0.37	1.27	0.39	0.99	0.36	0.74
Lithuania	2002	0.50	1.92	0.49	1.58	0.49	1.59	0.42	0.93
Mexico	2004	1.04	16.74	0.66	2.18	0.68	2.36	0.62	2.01
Mexico	2006	0.98	23.47	0.65	2.07	0.66	2.23	0.60	1.79
Mexico	2008	0.79	7.76	0.66	2.14	0.66	2.14	0.55	1.66
Moldova	2006	0.48	1.91	0.45	1.31	0.45	1.23	0.40	0.90
Mongolia	2007	0.63	26.04	0.42	1.28	0.42	1.28	0.40	1.21
Mozambique	2002	2.62	47.72	0.70	2.31	0.71	2.29	0.70	2.58
Myanmar	2006	0.47	17.82	0.34	1.04	0.34	0.94	0.30	0.79
Nepal	1995	1.36	35.66	0.48	1.59	0.48	1.80	0.48	2.07
Nicaragua	2005	0.45	2.58	0.42	1.41	0.42	1.44	0.36	0.93
Pakistan	2005	1.12	206.48	0.31	0.96	0.31	0.95	0.32	1.31
Panama	2008	8.01	65.83	0.56	1.86	0.60	1.98	0.74	13.12
Papua NG	1996	0.73	2.45	0.71	2.18	0.71	2.18	0.58	1.40
Paraguay	1997	0.55	2.46	0.50	1.59	0.51	1.58	0.44	1.11
Peru	2003	0.56	6.13	0.39	1.12	0.40	0.96	0.38	0.70
Philippines	2003	0.36	1.74	0.35	1.12	0.34	1.12	0.30	0.83
Sri Lanka	1999	1.42	47.55	0.37	1.01	0.37	1.30	0.39	2.34
Sudan	2009	0.87	7.21	0.70	2.33	0.71	2.18	0.61	1.37
Tajikistan	2007	0.39	2.39	0.36	1.08	0.37	1.19	0.32	0.93
Tanzania	2000	0.56	2.08	0.54	1.56	0.54	1.50	0.49	1.22
Tanzania	2007	5.23	41.80	1.36	3.17	2.36	5.91	1.96	5.23
Thailand	2011	0.50	2.76	0.45	1.43	0.45	1.41	0.40	0.96
Uganda	2005	0.58	2.78	0.54	1.67	0.53	1.50	0.46	1.07
Vietnam	1997	0.28	1.79	0.25	0.30	0.25	0.61	0.22	0.38
Vietnam	2006	0.31	1.74	0.28	0.89	0.28	0.63	0.25	0.44
Root Mean Square Difference from Untreated Data				**1.44**	**35.58**	**1.37**	**35.43**	**1.37**	**34.94**

Table 3: PoU for each data treatment method cross-tabbed with each decision-rule

		Function 1			Function 2			Function 3		
		Method 1 (%)	Method 2 (%)	Method 3 (%)	Method 1 (%)	Method 2 (%)	Method 3 (%)	Method 1 (%)	Method 2 (%)	Method 3 (%)
Albania	2005	6	5	2	9	9	6	9	9	6
Azerbaijan	2006	2	1	1	2	2	2	2	2	2
Bangladesh	2000	16	16	16	16	16	16	16	16	16
Bangladesh	2005	9	9	11	11	11	11	11	11	11
Bangladesh	2010	12	13	12	13	13	12	13	13	13
Bulgaria	2001	1	1	3	7	7	6	7	7	6
Burkina Faso	2003	28	28	22	24	24	19	24	24	19
Cambodia	2003	28	30	29	25	26	25	25	26	25
Cambodia	2009	16	21	23	15	19	20	15	19	20
Chad	2009	43	44	43	41	41	41	41	41	41
Congo, DR	2004	56	56	55	59	59	56	59	59	56
Egypt	1997	0	0	0	2	2	1	2	2	1
Ethiopia	1999	60	60	58	59	59	58	59	59	57
Georgia	2005	32	33	27	28	29	23	28	29	23
Guatemala	2006	27	27	26	24	24	23	24	24	23
Hungary	2002	0	0	0	2	2	1	2	2	1
India	2000	20	20	19	19	19	19	19	19	18
India	2004	23	22	21	21	22	20	21	22	20
Indonesia	2008	2	2	6	7	7	6	7	7	7
Iraq	2007	17	18	15	15	16	15	15	16	15
Cote d'Ivoire	2002	23	24	22	19	20	19	19	20	19
Lao PDR	2002	32	38	30	32	34	30	32	34	30
Lao PDR	2007	29	29	26	25	25	26	25	25	25
Lithuania	2002	0	0	0	4	4	3	4	4	3
Mexico	2004	6	7	1	8	8	6	8	8	6
Mexico	2006	3	4	0	7	7	5	7	7	5
Mexico	2008	25	25	23	22	22	20	22	22	20
Moldova	2006	15	15	11	14	14	12	14	14	12
Mongolia	2007	35	35	33	31	31	29	31	31	29
Mozambique	2002	42	42	41	39	39	38	39	39	38
Myanmar	2006	32	31	27	28	28	25	28	28	25
Nepal	1995	30	30	29	26	27	25	26	27	25
Nicaragua	2005	26	26	20	22	22	18	22	22	18
Pakistan	2005	15	15	14	15	15	14	15	15	14
Panama	2008	24	26	22	21	23	19	21	23	19
Papua NG	1996	27	27	23	24	24	20	24	24	20
Paraguay	1997	15	14	9	14	14	11	14	14	11
Peru	2003	26	25	21	22	23	21	22	23	21
Philippines	2003	23	23	17	20	20	17	20	20	17
Sri Lanka	1999	22	23	22	20	20	20	20	20	20
Sudan	2009	25	27	24	22	24	21	22	24	21
Tajikistan	2007	36	36	34	32	32	31	32	32	31
Tanzania	2000	42	42	41	39	39	37	39	39	37
Tanzania	2007	34	36	33	31	33	29	31	33	29
Thailand	2011	0	0	0	4	4	4	4	4	4
Uganda	2005	27	26	23	23	23	21	23	23	21
Vietnam	1997	30	32	30	30	32	30	30	31	30
Vietnam	2006	5	7	8	8	7	8	8	8	8

Table 4: Correlations of CVs for different methods for controlling for excess variability

	Empirical	Income Decile	Income Regression (equation 4)
Empirical	1.00		
Income Decile	0.29	1.00	
Income Regression (equation 4)	0.34	0.96	1.00

Figure 13: The Relative Price of Food indicator by region

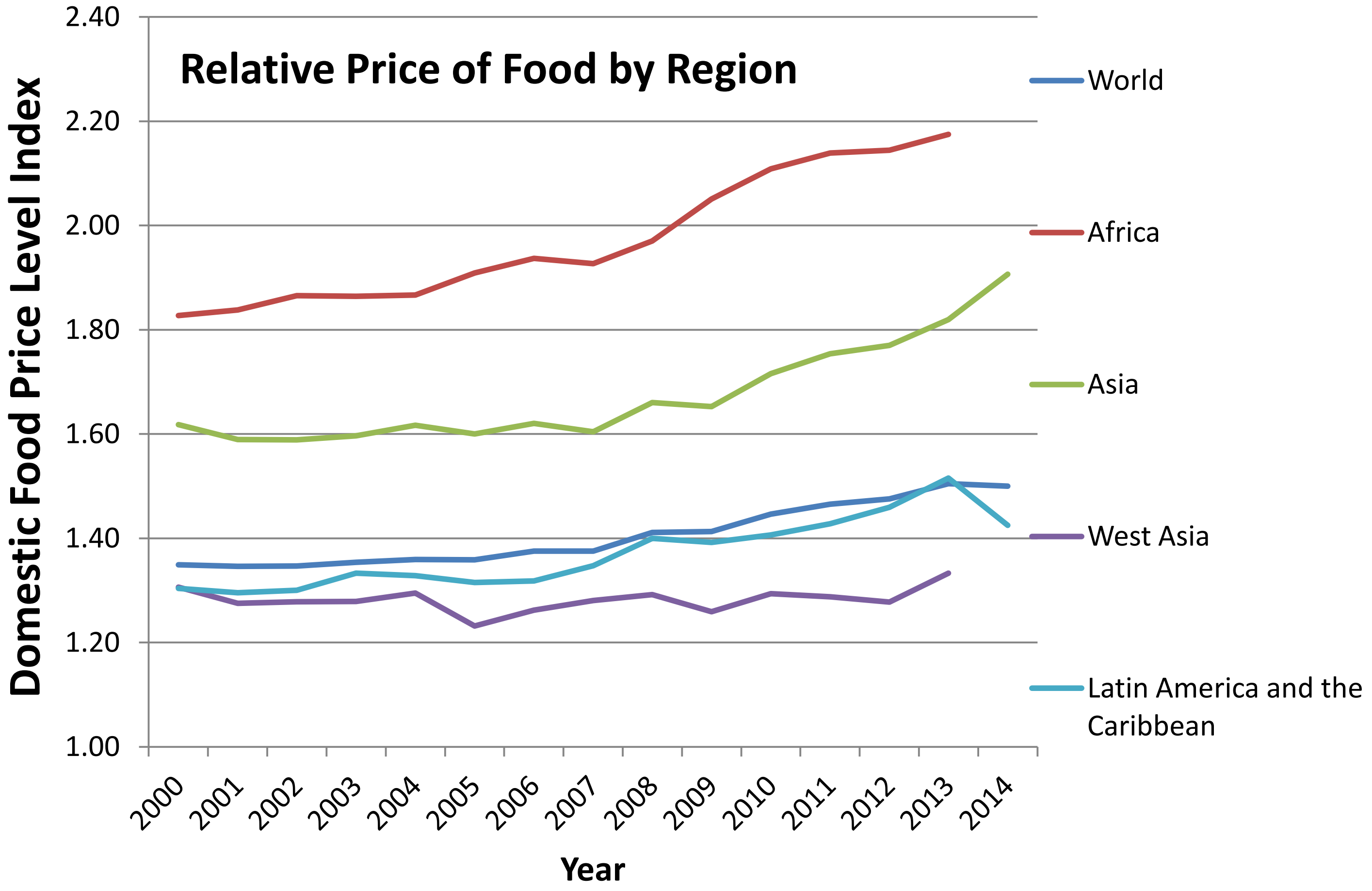


Table 5: Output Regression 1

Coefficients:	Estimate	Std. Error	Pr(>|t|)	Significance
Intercept	0.243	0.0172	< 2e-16	***
Gini	0.0392	0.0122	0.00250	**
log(GDP)	-0.0328	0.00964	0.00151	**
Region: Americas	-0.0172	0.0314	0.587	
Region: Asia	-0.0372	0.0214	0.0906	.
Region: Western Asia	-0.0555	0.0374	0.146	

Signif. codes: *** - 0.001 ** - 0.01 * - 0.05 . - 0.1

Residual standard error: 0.04047 on 40 degrees of freedom

Multiple R-squared: 0.6687, Adjusted R-squared: 0.6273

F-statistic: 16.15 on 5 and 40 DF, p-value: 1.06e-08

Note: Regional dummy variables are presented relative to Africa

Table 6: Output Regression 2

Coefficients:	Estimate	Std. Error	Pr(>|t|)	Significance
Intercept	0.282	0.0167	7.15e-16	***
Gini	0.0251	0.0102	0.0206	*
log(GDP)	0.0663	0.0241	0.0104	*
log(Food Price)	0.296	0.118	0.0182	*
Interaction	-0.171	0.0615	0.00980	**
Region: Americas	-0.163	0.0362	0.000116	***
Region: Asia	-0.0871	0.0213	0.000353	***
Region: Western Asia	-0.138	0.0401	0.00194	**

Signif. codes: *** - 0.001 ** - 0.01 * - 0.05 . - 0.1

Residual standard error: 0.02707 on 27 degrees of freedom

Multiple R-squared: 0.8423, Adjusted R-squared: 0.8015

F-statistic: 20.61 on 7 and 27 DF, p-value: 2.708e-09

Note: Regional dummy variables are presented relative to Africa

Table 7: Correlation matrices between independent variables in regressions

Survey Based	Gini	log(GDP)	log(Relative Price)
Gini	1.000		
log(GDP)	0.146	1.000	
log(Relative Price)	-0.151	-0.438	1.000

Country Updates	Gini	log(GDP)	log(Relative Price)
Gini	1.000		
log(GDP)	0.159	1.000	
log(Relative Price)	-0.285	-0.583	1.000

Appendix 1: Surveys and variables used in regression

Country	Year	CV│y	Gini	Notes	GDP	Relative Food Price	Region
Afghanistan	2007	0.18	30.3	Survey Based	938		Asia
Armenia	2004	0.05	25.7		3 646	1.82	Western Asia
Armenia	2011	0.11	33.7	Survey Based	6 891	1.87	Western Asia
Azerbaijan	2011	0.03	16.2	Survey Based	8 797		Western Asia
Azerbaijan	2006	0.07	17.5	2005	5 981	1.91	Western Asia
Bangladesh	2000	0.17	31.2		949	1.54	Asia
Bangladesh	2005	0.16	40.9		1 144	1.55	Asia
Bangladesh	2010	0.17	31.7		1 464	1.60	Asia
Cambodia	2003	0.20	41.3	2004	1 246	1.65	Asia
Cambodia	2009	0.18	31.4	2008	1 857	1.82	Asia
Chad	2009	0.31	60.7	Survey Based	1 655	2.62	Africa
Cote d'Ivoire	2002	0.32	44.2		1 774	2.03	Africa
Ecuador	2005	0.15	53.6		7 129	1.53	Americas
Ethiopia	1999	0.24	29.8	Survey Based	503	1.77	Africa
Guatemala	2006	0.23	56.2		4 188		Americas
Haiti	1999	0.45	59.2	2001	1 156		Americas
India	2004	0.15	32.4		2 074	1.54	Asia
India	2009	0.14	33.9		2 861	1.55	Asia
India	2000	0.16	32.0		1 745	1.59	Asia
Indonesia	2008	0.23	36.0	2009	3 581	1.77	Asia
Iraq	2007	0.19	30.9		2 951		Western Asia
Kazakhstan	2011	0.17	29.0	2009	11 568		Asia
Kazakhstan	2005	0.19	27.3		8 699	1.40	Asia
Lao PDR	2007	0.16	35.4		1 896	2.08	Asia
Lao PDR	2002	0.23	34.4		1 456	1.95	Asia
Mexico	2008	0.25	50.2		12 711	1.23	Americas
Mexico	2004	0.29	50.6		11 807	1.17	Americas
Mexico	2006	0.25	49.6		12 462	1.20	Americas
Mongolia	2007	0.21	35.8		3 362	1.88	Asia
Mozambique	2002	0.31	46.4		574	1.91	Africa
Nicaragua	2005	0.26	51.1		3 013		Americas
Pakistan	2005	0.14	32.5		2 154	1.88	Asia
Peru	2003	0.18	52.0		5 797	1.54	Americas
Philippines	2003	0.23	44.5		2 826	1.60	Asia
Sri Lanka	1999	0.19	31.8		2 842	1.72	Asia
Sudan	2009	0.22	34.4		1 765		Africa
Tajikistan	2005	0.25	33.6	2004	1 423		Asia
Tajikistan	2007	0.23	32.6		1 573		Asia
Tanzania	2000	0.25	34.6		868	1.83	Africa
Tanzania	2007	0.31	37.6		1 156	1.96	Africa
Thailand	2011	0.11	39.3	2009	7 972	1.90	Asia
Timor Leste	2001	0.26	43.7	Survey Based	1 201		Asia
Uganda	2002	0.30	43.6		826	1.50	Africa
Uganda	2005	0.24	42.2		902	1.68	Africa
Vietnam	2006	0.19	37.7		2 490	1.76	Asia
Zambia	2002	0.32	41.6		1 070	1.69	Africa

Notes: "Survey" means the Gini coefficient of income was directly calculated from the survey. A year in the notes column means that a slightly different year was used from the Gini coefficient dataset.

Appendix 2(a): Series of Gini coefficients used for updates

Country	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Bolivia	59.1	54.9	56.1	54.5		54.6	54.2	51.9	51.5	46.5		43.6
Brazil									54.2	53.6		52.7
Cameroon		40.4	44.1					39.0				
Cape Verde		72.4	73.5									
Central African Rep				43.3					56.2			
Sri Lanka			39.8					40.3				
Colombia	57.2	56.5	57.4	53.2		54.8	58.7		55.5	55.2		53.5
Ecuador						53.6	52.9	53.9	50.2	48.8	48.9	45.8
Georgia	41.1	39.0	38.0	39.3	40.6	39.9	39.4	39.4	39.0			
Ghana						42.4	42.8					
Guinea				42.9				39.6				
Indonesia										36.0	32.3	
Jordan			38.6	38.9					33.8		35.4	
Kenya						47.1		29.9				
Korea Rep.				33.6		34.1	32.6		34.7			
Laos								35.4	36.7			
Lesotho			52.0	52.5								
Madagascar		45.9				69.7					44.1	
Malawi					39.0	38.6					45.2	
Malaysia					37.7			46.0		46.8		
Mali		38.9					38.6				33.0	
Mauritania	38.9				41.3				40.5			
Mexico									50.2		47.5	
Morocco		40.3						40.7				
Mozambique			46.4	47.1		41.3			45.6			
Nepal				43.2							32.8	
Niger						43.4		37.3	34.6			
Nigeria				41.8	40.9						46.8	
Paraguay		55.8	56.4	56.7	53.8	52.9	54.5	54.2	52.1	50.7	52.2	54.3
Peru				52.0		49.8	49.6	50.5	48.0	48.0	47.2	45.7
Philippines				44.5			44.1			39.2		
Rwanda	51.5						53.1					50.8
Senegal		46.3				38.9						
Sierra Leone				38.1		42.2						
South Africa	57.3						67.4			63.1		
Turkey	40.1		42.7		42.7	41.8	40.3	39.3				
Uganda						42.2	42.6			42.6	42.9	
Burkina Faso				37.6						39.8		
Uruguay	44.4	46.2	46.6	46.2		45.9	47.2	47.6	46.3	46.3	45.3	43.4
Venezuela	44.1	46.4	47.4	46.1	45.3	47.5	43.4	41.5	40.3	40.0	38.4	38.8
Vietnam							37.7		35.6			
Ethiopia	27.9				28.9	29.8						
Yemen						37.3			37.4			
Zambia			41.6	42.1	50.3		54.6				57.4	

Appendix 2(b): Series of GDP used for updates: GDP per capita, PPP (constant 2005 international $)

Country	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Bolivia	4 330	4 315	4 335	4 367		4 578	4 716	4 850	5 066	5 152		5 462
Brazil									13 338	13 176		14 301
Cameroon		2 339	2 371					2 440				
Cape Verde		3 559	3 678									
Central African Rep				674					797			
Sri Lanka			5 112					6 587				
Colombia	8 433	8 434	8 508	8 702		9 306	9 781		10 512	10 534		11 364
Ecuador						8 359	8 574	8 611	9 005	8 906	9 019	9 569
Georgia	3 276	3 459	3 673	4 105	4 356	4 727	5 128	5 773	5 913			
Ghana						2 521	2 613					
Guinea				1 194				1 203				
Indonesia										7 661	8 030	
Jordan			8 160	8 292					10 897		11 254	
Kenya						1 857		2 003				
Korea Rep.				22 203		24 008	25 129		26 701			
Laos								3 298	3 483			
Lesotho			1 683	1 749								
Madagascar		1 481				1 388					1 387	
Malawi					716	716					736	
Malaysia					17 335			19 386		19 312		
Mali		1 386					1 564				1 668	
Mauritania	2 263				2 308				2 807			
Mexico									15 704		15 335	
Morocco		4 676						5 784				
Mozambique			632	652		728			819			
Nepal				1 635							2 011	
Niger						788		799	843			
Nigeria				3 165	4 125						5 148	
Paraguay		6 060	5 938	6 074	6 199	6 212	6 391	6 616	6 912	6 521	7 247	7 431
Peru				7 132		7 811	8 323	8 966	9 739	9 724	10 460	11 049
Philippines				4 465			4 967			5 304		
Rwanda	746						1 026					1 313
Senegal		1 953				2 106						
Sierra Leone				1 133		1 158						
South Africa	9 488						11 023			11 410		
Turkey	13 090		12 733		14 258	15 252	16 093	16 634				
Uganda						1 033	1 106			1 264	1 294	
Burkina Faso				1 179						1 352		
Uruguay	12 415	11 916	10 993	11 088		12 512	13 005	13 822	14 768	15 047	16 338	17 345
Venezuela	14 461	14 673	13 129	11 894	13 820	14 981	16 178	17 298	17 911	17 056	16 536	16 960
Vietnam							3 687		4 085			
Ethiopia	612				666	724						
Yemen						4 271			4 380			
Zambia			2 170	2 225	2 286		2 427				2 779	

Appendix 2(c): Series of Relative Price used for updates

Country	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Bolivia	1.73	1.71	1.67	1.68		1.71	1.75	1.81	1.67	1.57		1.64
Brazil									1.28	1.27		1.31
Cameroon		1.96	2.00					1.95				
Cape Verde		1.58	1.50									
Central African Rep				2.25					2.24			
Sri Lanka			1.75					1.80				
Colombia	1.60	1.59	1.61	1.62		1.63	1.67		1.72	1.68		1.68
Ecuador						1.53	1.58	1.57	1.60	1.58	1.61	1.64
Georgia	1.51	1.54	1.56	1.60	1.62	1.70	1.76	1.75	1.74			
Ghana						2.37	2.35					
Guinea				2.19				2.66				
Indonesia										1.78	1.87	
Jordan			1.17	1.18					1.29		1.32	
Kenya						1.90		2.05				
Korea Rep.				1.82		1.87	1.86		1.82			
Laos								2.08	2.10			
Lesotho			2.16	2.05								
Madagascar		2.00				2.16					2.00	
Malawi					2.08	2.13					1.99	
Malaysia					1.48			1.52		1.56		
Mali		2.02					1.99				2.00	
Mauritania	2.01				2.10				2.20			
Mexico									1.23		1.25	
Morocco		1.56						1.60				
Mozambique			1.91	1.93		1.90			2.01			
Nepal				1.50							1.63	
Niger						2.05		2.16	2.24			
Nigeria				2.38	2.34						2.42	
Paraguay		1.32	1.31	1.40	1.43	1.42	1.51	1.62	1.67	1.60	1.68	1.75
Peru				1.54		1.56	1.58	1.58	1.60	1.59	1.62	1.63
Philippines				1.60			1.59			1.62		
Rwanda	1.52						1.74					1.70
Senegal		2.04				2.07						
Sierra Leone				2.29		2.40						
South Africa	1.29						1.43			1.47		
Turkey	1.41		1.41		1.40	1.48	1.63	1.76				
Uganda						1.68	1.74			1.90	1.88	
Burkina Faso				1.86						1.97		
Uruguay	1.23	1.21	1.21	1.23		1.26	1.27	1.33	1.38	1.34	1.36	1.38
Venezuela	1.27	1.31	1.37	1.44	1.57	1.66	1.77	1.87	2.04	2.01	2.10	2.15
Vietnam							1.75		1.66			
Ethiopia	1.74				1.75	1.80						
Yemen						1.54			1.71			
Zambia			1.69	1.70	1.66		1.63				1.47	

References:

1. "United Nations Millennium Development Goals." http://www.un.org/millenniumgoals/
2. "The World Food Summit Goal and the Millennium Development Goals." *Twenty-seventh Session of Committee on World Food Security* (Rome, 28 May - 1 June 2001).
3. Cafiero, Carlo. "Advances in Hunger Measurement." *International Scientific Symposium on Food and Nutrition Security Information: from Valid Measurement to Effective Decision Making*, Rome 2012.
4. Azzalini, Adelchi. "A class of distributions which includes the normal ones." *Scandinavian journal of statistics* (1985): 171-178.
5. Azzalini, Adelchi, Thomas Del Cappello, and Samuel Kotz. "Log-skew-normal and log-skew-t distributions as models for family income data." *Journal of Income Distribution* 11.3 (2002).
6. FAO. 2011. Global food losses and food waste: Extent, causes and prevention, by D. Gustavsson, C. Cederberg, U. Sonesson, R. van Otterdijk and A. Meybeck. Rome.
7. Human energy requirements: Report of a Joint FAO/WHO/UNU Expert Consultation. Rome, 17-24 October 2001. United Nations University, World Health Organization and Food and Agriculture Organization of the United Nations.
8. Kim, Tae-Hwan, and Halbert White. "On more robust estimation of skewness and kurtosis." *Finance Research Letters* 1.1 (2004): 56-73.
9. Bluman, Allan G. *Elementary Statistics: A Step by Step Approach*. Boston: McGraw-Hill, 2004.
10. Peter Rousseeuw, Christophe Croux, Valentin Todorov, Andreas Ruckstuhl, Matias Salibian Barrera, Tobias Verbeke, Manuel Koller, Martin Maechler (2013). robustbase: Basic Robust Statistics. R package version 0.9-10. URL http://CRAN.R-project.org/package=robustbase
11. Witten, I. H., and Eibe Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Amsterdam: Morgan Kaufman, 2005.
12. Pardoe, Iain. *Applied Regression Modeling*. Hoboken, NJ: Wiley, 2012.
13. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.
14. Sibrian, Ricardo. "Deriving Food Security Information from National Household Budges Surveys: Experiences, Achievements, Challenges." *Food and Agriculture Organization of the United Nation*
15. *Proceedings: Measurement and Assessment of Food Deprivation and Undernutrition: International Scientific Symposium, Rome, 26-28 June 2002*. Rome: Food and Agriculture Organization of the United Nations, 2003.
16. Smith, Lisa C. "Can FAO's measure of chronic undernourishment be strengthened?." *Food Policy* 23.5 (1998): 425-445.
17. *All the Ginis* database. *World Bank.* http://econ.worldbank.org/projects/inequality
18. "2005 International Comparison Program Tables of Final Results." *World Bank*
19. "FAOSTAT." Food and Agriculture Organization of the United Nations. http://faostat.fao.org/
20. "Food Security Indicators." *ESS Website Food and Agriculture Organization of the United Nations.* http://www.fao.org/economic/ess/ess-fs/ess-fadata/en/
21. Solt, Frederick. "Standardizing the world income inequality database." *Social Science Quarterly* 90.2 (2009): 231-242.
22. "World Population Ageing 2013." United Nations, New York 2013.
23. "United Nations Population Division | Department of Economic and Social Affairs." http://www.un.org/en/development/desa/population/
24. Deaton, Angus, and Jean Drèze. "Food and nutrition in India: facts and interpretations." *Economic and political weekly* (2009): 42-65.
25. Svedberg, Peter. "841 million undernourished?." *World Development* 27.12 (1999): 2081-2098.

Contact:
Nathan Wanner
Statistics Division (ESS)
The Food and Agriculture Organization of the United Nations
Viale delle Terme di Caracalla
00153 Rome, Italy

www.fao.org/economic/ess/ess-publications/workingpapers/

FAO-statistics@fao.org

I4046E/1/09.14