



IARIW 2025

# IARIW 2025

Thursday, October 2 & Friday, October 3

## **Bottom-Coding Threshold in the Poverty and Inequality Platform**

Nishant Yonzan

(World Bank)

Dean Jolliffe

(World Bank)

Christoph Lakner

(World Bank)

Paper prepared for the IARIW–World Bank–UEB/VNU Conference on “Improving Well-being Measurement in Data-challenged Environments in Developing Countries for Better Evidence-based Policies” October 2-3, 2025

Session 2B Data and Methods

Time Slot: Thursday, October 2, 1:15- 3:15 PM

# Bottom-coding for the measurement of global poverty and inequality

Nishant Yonzan  Minh C. Nguyen  Christoph Lakner  Aart Kraay

 Dean Jolliffe  Haoyu Wu  Gabriel L. Ibarra<sup>1</sup>

September 12, 2025

**(Preliminary draft prepared for the IARIW-World Bank-UEB/VNU Special Conference “Improving Well-being Measurement in Data-Challenged Environments in Developing Countries for Better Evidence-based Policies”. Please do not cite.)**

**Abstract:** Sensitivity to low incomes is a desirable feature of any distribution sensitive measure. However, some welfare measures cannot incorporate negative or zero incomes. Furthermore, small positive values can have an extreme influence on distribution sensitive indices. Hence, income or consumption is often coded at the bottom at some strictly positive value. However, it is unclear where the bottom-coding threshold should be. The World Bank’s Poverty and Inequality Platform has used a threshold at zero or a small positive number depending on the welfare measure estimated, while Luxembourg Income Study uses a bottom threshold that is three times the inter-quartile range below the first quartile of log incomes in each survey to report their standard inequality statistics. This paper explores various options to bottom code consumption and income distributions for cross-country comparison. Using close to 2,000 consumption and income surveys from across the world and employing various methodologies, this paper finds that \$0.25 per person per day in 2017 PPP is a suitable threshold to bottom code the distribution.

*Keywords: Bottom code, consumption floor, measurement errors, distribution-sensitive index, parametric tail*

*JEL Classifications: C1, D60, I32*

---

<sup>1</sup> All authors are with the World Bank. Corresponding author: Nishant Yonzan ([nyonzan@worldbank.org](mailto:nyonzan@worldbank.org)). We would like to thank the Global Poverty Working Group of the World Bank for helpful comments. The names of the authors are in random order, using the author randomization tool of the American Economic Association, which is archived here. We acknowledge financial support from the UK government through the Data and Evidence for Tackling Extreme Poverty (DEEP) Research Program. The views expressed here are the authors’, and do not necessarily reflect the official views of the World Bank, its Executive Directors, or the countries they represent.

## 1. Introduction

Sensitivity to low incomes is a desirable feature of any distribution sensitive measure. However, some welfare measures cannot incorporate negative or zero incomes. Furthermore, very small positive values can have an extreme influence on distribution sensitive indices (Cowell and Victoria-Frazer 2006; Cowell and Flachaire, 2007). Hence, income or consumption is often coded at some strictly positive minimum value.<sup>2</sup> However, there is no clear guidance in the literature for a threshold to use as a bottom code, nor is there a discussion on the differences between using a threshold for consumption or a threshold for income distribution.

In addition to the sensitivity of the negative, zero, and small values to poverty and inequality indices, the data at the bottom of the income and consumption distributions are known to have measurement issues due to transient factors and measurement errors (Ravallion, 2016). For example, quite a few observations in surveys from high-income countries have income between zero and \$0.005 per day (2017 PPP\$) or as many as 15 consumption surveys in the World Bank's Global Monitoring Database have observations with zero consumption. For consumption surveys, zero and very low reported consumption is likely to be the result of measurement error given that there is a biological minimum consumption level required to sustain life. For income surveys, very low, zero, and even negative incomes are plausible observations as individuals can finance consumption by drawing down savings, making it difficult to argue that low observed values are due to measurement error. Even in the case of income, however, the minimum threshold for consumption could be a satisfactory threshold to bottom code for measuring poverty and inequality indicators, recognizing that the consumption levels of those individuals with low income are unlikely to be lower and could well be considerably higher.<sup>3</sup>

Given the inadequacy of the current ad-hoc adjustments and the potential for measurement error, there is a need to define a threshold to bottom code when reporting poverty and inequality measures.<sup>4</sup> Traditional measures of poverty, such as the FGT headcount measure, which only counts the number of people below a poverty threshold are unaffected or less affected by the small income values. On the other hand, distribution sensitive measures such as the Prosperity

---

<sup>2</sup> For example, the Luxembourg Income Study bottom- (and top-) codes the distribution of log income at three times the inter-quartile range below (above) the first (third) quartile when reporting inequality measures (Neugschwender, 2020). Until recently, all indicators in the Poverty and Inequality Platform (PIP) were calculated using a distribution of income or consumption that is truncated at a value equal to zero (i.e., observations with a negative value are dropped). In addition to this, ad-hoc adjustments were made for those indicators that cannot accept zero value; for instance, in the case of the mean log deviation observations with a value of zero are replaced with a small positive value, while those observations are dropped in the case of the Watts index.

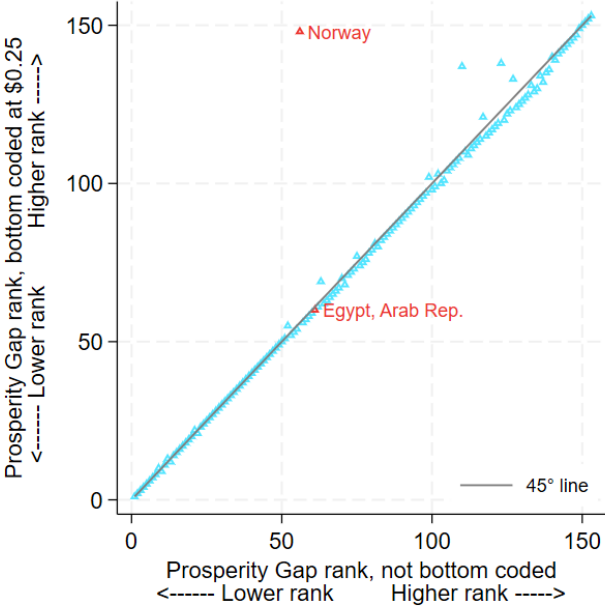
<sup>3</sup> We have examined 20 income surveys in the Global Monitoring Database in which measured incomes are zero in the lowest percentiles of the income distribution, *and* a consumption survey is available in the same year. These countries are all upper middle-income or high-income countries in Eastern Europe. Pooling across all surveys, median consumption over those percentiles for which reported incomes are zero is \$2.64 per day, with the lowest consumption percentile being \$0.84 per day.

<sup>4</sup> Two adjustments are possible: (a) truncate the distribution below the threshold, or (b) assign the values below the threshold the threshold value. In the paper, bottom-code refer to the latter. However, for reasons explained further below, negative values are dropped from the distribution.

Gap are much more affected by small values (Kraay et al. 2023). This paper justifies a threshold where a bottom code can be consistently applied with minimal impact across all poverty and inequality measures.

To highlight the concern, Figure 1 reports the impact of bottom coding on the Prosperity Gap. The Prosperity Gap is the average factor by which incomes must be multiplied for a society on average to reach \$25 per day in 2017 PPPs. Since the Prosperity Gap is a factor, it is strongly sensitive to small values of income or consumption; the factor is double for a person with half the income, quadruple for a person with one-fourth the income, and so on (Kraay et al. 2023). Using the latest survey available for the country, the figure shows the rank-rank correlation of the Prosperity Gap estimated using a distribution that is bottom-coded at \$0.25 per day (vertical axis) and another that is not bottom coded (horizontal axis). Countries are ranked from the least prosperous (rank of 1) to the most prosperous. Most observations are close to the 45-degree line, suggesting limited re-ranking. However, there are some outliers. Not bottom coding would mean that Norway, for example, would be wrongly classified as less prosperous than Egypt, a country with the average household income more than 8-times lower, due to a few very small income observations. Bottom-coding addresses these data issues at the country-level, while it has a minimum impact on aggregate results.

Figure 1: The rank-rank correlation of the Prosperity Gap between bottom coding at \$0.25/day and not bottom coding



Note: Each observation is the latest (post 2000) survey for 170 countries ranked from the least prosperous (rank of 1) to the most prosperous. On the vertical axis countries are ranked after bottom coding the distribution at \$0.25 per person per day in 2017 PPPs. Observations with zero values are not included in both samples as they are dropped in the case with no bottom code.

We propose a threshold of \$0.25 per person per day (in 2017 PPPs USD) to bottom code consumption and income distributions. This would mean converting all non-negative values below the threshold to the threshold value, in other words winsorizing values below \$0.25 per day. The issue of negative income is beyond the scope of the work in this paper and thus we drop the negative incomes for this study. There is evidence (Hlasny, Ceriani, Verme 2022) to suggest that households with negative incomes are non-poor households, so it is unclear if assigning them a very low-income value is appropriate. The \$0.25 per day threshold is almost equivalent to the minimum of the least-cost energy sufficient diet; it is around the typical floor estimated using various statistical approaches; it is a threshold where measurement errors start becoming less obvious; and it is a threshold where the contribution of the bottom of the distribution is fairly in line with a well-behaved distribution. While a survey-specific bottom code would be ideal, there are costs to include multiple bottom codes when working with spatial and temporal comparisons, and, hence, a unique value is proposed for reporting of global welfare.

In what follows, Section 2 outlines the various methods used for estimating a threshold to code at the bottom of consumption and income distributions, Section 3 discusses the data, Section 4 outlines the findings, Section 5 summarizes the various approaches and reports the impact of the thresholds on various poverty and inequality measures, and Section 6 concludes.

## 2. Methodology

We use various statistical methods to estimate a threshold to bottom code the distribution of consumption. Since consumption distribution has a natural minimum, we will estimate the thresholds to bottom code primarily on the consumption surveys and report the impact of the threshold for inequality estimates in income surveys. Where applicable, we will report floors for income surveys. First, following Ravallion (2016), we estimate a “permanent consumption floor” (see also Jolliffe et al. 2023; Beduk, Jara, Ravallion, *forthcoming*). Second, we estimate a minimum value in the distribution by fitting various parametric distributions – beta distribution, three-parameter log normal, and Dagum III – considered appropriate for income and consumption distributions in the literature (see Dagum 1980; Kleiber 2007, 2008). This approach is in line with the principle of fitting a parametric distribution (such as the Pareto) at the top of the income distribution. In our case, our focus is at the bottom of the distribution (Cowell and Vicotria-Feser 1996; Cowell and Flachaire 2007). Third, we utilize outlier identification methodology, such as that adopted by the Luxembourg Income Study (LIS) database (Neugschwender 2020) to estimate a value to bottom code the distribution. Finally, we provide evidence of a minimum consumption from the energy sufficient diet literature (Herforth 2022).

In addition, we provide sensitivity checks by exploring the sensitivity of poverty and inequality measures to the thresholds derived above. These are presented in Section 5.

## 2.1. Permanent consumption floor using the Ravallion (2016) method

Ravallion (2016) argues that the strict lower bound of observed consumption or income in survey data would be “unsatisfactory as the permanent floor as observed minimum values [in survey data] are subject to transient factors, and measurement errors.” Instead, he argues that “the observed current consumption equals the floor for permanent consumption at some point within a range of low observed consumption”, and that the probability of the observed consumption value being the consumption floor declines as the observed consumption increases.

In other words, the permanent minimum consumption, or consumption floor, is given by the function:  $E[y^{min}|x] = \sum_{i=1}^n \phi(x_i)x_i$ , where  $y$  is permanent consumption,  $x$  is observed consumption, and  $\phi(\cdot)$  is the relative weight assigned to each observed value. Furthermore, with the assumption that the probability of finding a floor that is above the poverty line,  $z$ , is zero, a permanent consumption floor ( $y_f$ ) can be estimated by

$$y_f \equiv E_\alpha[y^{min}|x] = z \times \left(1 - \frac{P_{\alpha+1}}{P_\alpha}\right) \quad (1)$$

where  $P$  is the Foster-Greer-Thorbeck (FGT) poverty index (Foster et al. 1984) and  $\alpha \geq 1$  is the degree of curvature. For instance, when  $\alpha = 1$ , linear weights are assumed and  $P_\alpha = P_1$  is the poverty gap index, and  $P_{\alpha+1} = P_2$  is the squared poverty gap index. The higher the degree of curvature, the larger the weight is assigned to the values at the bottom of the distribution compared to those near the poverty line. The parameters to choose are the poverty threshold ( $z$ ) and the weights or curvature ( $\alpha$ ). Ravallion (2016) uses the International Poverty Line (IPL) as the threshold with  $\alpha = \{1,2\}$ .

In short, a permanent consumption floor is a weighted average consumption of the observations below a certain threshold. It is important to highlight that the object of this paper is to identify a suitable minimum value to bottom code the distribution in order to calculate poverty and inequality indicators. Identifying the permanent consumption might tell us something about the ceiling for the bottom coding value.

## 2.2. Consumption floor using various parametric distributions

### 2.2.1. Beta distribution

The approach outlined in Ravallion (2016) justifies the use of a weighted average at the bottom of the distribution by emphasizing measurement concerns at the bottom of the distribution. The approach is, however, not clear on the distributional assumption underlying the probability weights,  $\alpha$ , nor the choice of the poverty threshold; Ravallion suggests anchoring it to the International Poverty Line. Beduk et al. (forthcoming), using longitudinal data for Great Britain, argue that using the income or consumption of the person at the 10<sup>th</sup> percentile as the poverty

threshold along with  $\alpha = 1$  is a reasonable option to estimate a permanent consumption floor among various choices.

An alternate approach is to fit a known parametric distribution at the bottom. This is, in principle, like fitting a Pareto distribution to the top of the income distribution to correct for the missing data at the top (see Cowell and Victoria-Feser 1996; Cowell and Flachaire 2007). This abstracts away from the assumptions in Ravallion (2016) and at the same time uses an approach that is as convenient to apply to the data – i.e., using poverty threshold and the FGT indices. Suppose, instead, we assume that consumption levels between the consumption floor and  $z$  follow a shifted and rescaled Beta distribution, i.e.,  $y = y_f + (z - y_f)x$  with  $x \sim \text{Beta}(\eta, \beta)$ . As long as  $z$  is below the modal value of the overall density of the distribution, it is reasonable to impose the further restrictions that  $\eta > 1$  and  $\beta = 1$ . Together, this ensures that the density of  $y$  is monotonically increasing over the interval  $[y_f, z]$ .

Using moment conditions of the Beta distribution and relating it with the summary poverty statistics as Ravallion (2016), the floor can be estimated by

$$y_f = z \left( 1 + \frac{P_0 P_1 P_2}{P_0^2 P_2 - 2P_0 P_1^2} \right) \quad (2)$$

where  $P_\alpha$  is the FGT poverty index as in equation 1. See Appendix A for a detailed derivation of equation 2.

It is important to consider that while Ravallion (2016) attempts to find a “permanent” consumption floor, the  $y_f$  estimated by equation 2 is the minimum value in the survey using a parametric fit. Hence, the floors estimated using the Beta distribution and other parametric methods described below is not directly comparable to the permanent consumption floor, which is an average of values at the bottom of the distribution. Indeed, with the Beta distribution approach, there is no restriction that the  $y_f$  is non-negative.

### 2.2.2. *The three-parameter log normal distribution*

The next two parametric distributions fit the functional form on the entire distribution and not just the bottom tail. Thus, the focus is not only on the bottom as with the previous cases. However, both distributions are suited for fitting income and consumption distributions and both have a defined floor parameter. Additionally, as with the Beta distribution and unlike Ravallion (2016), the estimated minimum values from these parametric distributions are fitted minimum values rather than an estimate of a permanent consumption floor as in Ravallion (2016). Both the three-parameter log normal and the Dagum III distributions have a theoretical positive floor that makes these distributions appealing.

The three-parameter log normal (TPLN) distribution is a shifted log-normal distribution that includes a threshold parameter in addition to the well-known two parameter lognormal

distribution (see Aristizabal 2012; Cohen and Whitten 1980; Cohen 1951; Wicksell 1917). The TPLN is appealing as the threshold parameter equates to the floor of the distribution.

The density function of the TPLN can be written as

$$f(x; \gamma, \mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{1}{2}}(x - \gamma)^{-1} \exp\left\{-\frac{[\ln(x - \gamma) - \mu]^2}{2\sigma^2}\right\},$$

where,  $\sigma > 0$ ,  $-\infty < \mu < \infty$ , and  $\gamma < x < \infty$ .  $\mu$  is the mean,  $\sigma^2$  is the variance, and  $\gamma$  is the threshold parameter of interest. Following Cohen & Whitten (1980), we estimate  $\gamma$  using local maximum likelihood estimator (LMLE, see Appendix B) and method of moments. The formula for method of moments estimator (ME) in equation 3 below is replicated from Cohen & Whitten (1980).

$$\hat{\gamma} = \bar{x} - s \times (\omega - 1)^{-\frac{1}{2}} \quad (3)$$

where,  $\bar{x}$  is the sample mean,  $s$  is the sample standard deviation,  $\omega$  is the real root from  $\omega^3 + 3\omega^2 - (a^2 + 4) = 0$ , and  $a$  is the third standard moment of the sample. Cohen & Whitten (1980) recommend using a version of the ME when the third moment of the distribution is less than one and using a version of the LMLE otherwise.

### 2.2.3. The Dagum III distribution

Dagum distributions have been used to fit the distribution of income and wealth (see Dagum 1980; Kleiber 2007; Kleiber 2008). The Dagum distribution, and others such as the Singh-Maddala and Burr III/XII, are a family of distribution that can be nested within the four-parameter generalized beta distribution. The Dagum distribution in particular is well suited for studying income distributions as it improves on the log-normal distribution by allowing for a thick right tail and improves on the Pareto distribution by permitting an interior mode (Kleiber 2008).

The Dagum III is the variant that supports strictly positive values and has a positive minimum value. The density function of the Dagum III distribution is as follows

$$f(x; a, b, p) = ab^a p(x - x_0)^{-a-1} (1 + b^{a(x-x_0)^{-a}})^{-p-1},$$

where,  $a, b, p > 0$  and  $0 < x_0 \leq x < \infty$ . We use maximum likelihood to estimate the minimum value,  $x_0$ .<sup>5</sup>

---

<sup>5</sup> The floor,  $x_0$ , can also be estimated using  $\left\{b \left[ \left(1 - \frac{1}{a}\right)^{\frac{1}{p}} - 1 \right]\right\}^{-\left(\frac{1}{a}\right)}$  once the  $a$ ,  $b$ , and  $p$  have been estimated (see Kleiber 2008 for detail). The maximum likelihood formulation is reported in Appendix C.

### 2.3. Floor using outlier detection method

Finally, if measurement issues are a concern, outlier detection methods can be helpful in identifying such observation. The traditional approach for this kind of outlier detection is to induce normality in the distribution then define a threshold in a distribution of standard normal using z-scores. Typically, the method boils down to defining the floor that is at some distance from a moment of the distribution as in

$$y_f = \mu - z_\alpha \times \sigma$$

where the threshold,  $y_f$  is  $z$  standard deviations,  $\sigma$ , below the mean,  $\mu$ .

One such method is applied to the standard inequality estimates reported in the Luxembourg Income Study (LIS) database (Nurembourg 2020). LIS defines a floor (and also a ceiling) at three times the inter quartile range ( $IQR$ ) below (above) the first quartile (third quartile) of the distribution of log income. We supplement this by reporting a range of values considering floors that is two- to four-times the  $IQR$  below the first quartile. Note however that the floors that are derived using this method is a function of the  $IQR$ . Hence, all else equal, the larger the  $IQR$  the lower the estimated floor. By construction, distributions that are more dispersed have lower floors. The floors can be expressed as

$$y_f = \exp(Q_1 - k \times (Q_3 - Q_1)) \quad (3)$$

where  $Q_q$  is the  $q^{th}$  quartile of the log of consumption and  $k \in \{2,3,4\}$ .

## 3. Data

The primary data source for this study is the Global Monitoring Database (GMD), a comprehensive repository of survey microdata maintained by the World Bank. The GMD consists of nationally representative household surveys conducted by national statistical offices in collaboration with the World Bank. These surveys are subsequently compiled, processed, and harmonized to ensure consistency and comparability across different countries and time periods. The GMD hosts data for 122 countries covering the period 1981 to 2023. For additional 34 countries we use microdata from European Union's Statistics on Income and Living Condition surveys (EU-SILC) covering the period 2004 to 2023.<sup>6</sup>

The household survey datasets in the GMD provide extensive information on household income, consumption, and various harmonized demographic variables, such as household size, age,

---

<sup>6</sup> GMD also hosts an additional 165 surveys that are based on grouped data, 13 surveys based on imputed data, 308 surveys based on binned data from Luxembourg Income Study database. Only surveys with microdata (household-record level) are included in this study.

gender, and the rural/urban classification of households. This harmonization allows for the analysis of poverty levels and trends, as well as other key sociodemographic attributes, facilitating reasonable comparisons both across and within countries over time. The GMD serves as the primary data source for the World Bank’s poverty and inequality estimates, as published in the Poverty and Inequality Platform (PIP) and subnational poverty portal, and for the calculation of the World Bank’s multidimensional poverty measure (MPM).

The data used in this analysis contains 1,826 household surveys (country-year observation) conducted in about 156 economies across 40 years (1981 to 2023). Among them, 795 surveys carry consumption welfare, 1,003 surveys carry income welfare, and 14 carry both welfare aggregates. Table 1 provides an overview of the distribution of surveys conducted across different regions, distinguishing between consumption welfare and income welfare surveys. Notably, Europe & Central Asia (ECA) leads with a total of 921 surveys, split between 372 on consumption welfare and 549 on income welfare, representing the highest number in each category. In contrast, most surveys in Latin America & the Caribbean (LAC) focus on income welfare, while the majority of surveys in Sub-Saharan Africa (SSA) are on consumption welfare. Despite having the most countries, ECA maintains a high average of 20 surveys per country, which is five times higher than the lowest average of 4 surveys per country in SSA. Overall, the table highlights significant regional variations in both the distribution of consumption and income welfare surveys, as well as the frequency of surveys per country, reflecting the different priorities and data collection efforts across regions.

Table 1: Distribution of survey data by region

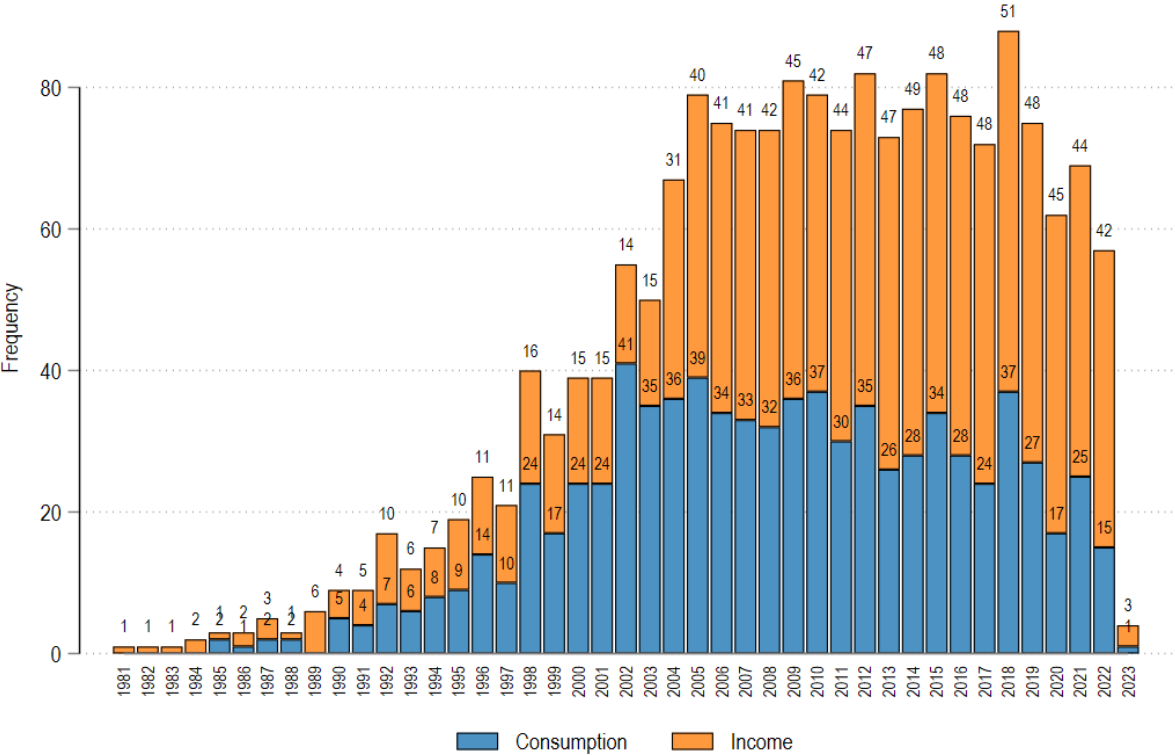
Regions	Number of surveys		Number of countries	Average survey per country
	Consumption welfare	Income welfare		
East Asia & Pacific	120	17	20	7
Europe & Central Asia	372	549	47	20
Latin America & the Caribbean	18	434	25	18
Middle East & North Africa	56	15	12	6
South Asia	45		7	6
Sub-Saharan Africa	198	2	45	4
<b>Total</b>	<b>809</b>	<b>1017</b>	<b>156</b>	<b>12</b>

Note: Groupings are primarily based on the regions used for [administrative purposes](#) by the World Bank.

Over the decades, the frequency of household surveys received has exhibited clear trends, reflecting shifts in data availability and the focus of economic analysis. From 1981 to early 1990s, the number of surveys received was relatively low, with only one or two per year. However, from

the mid-1990s onward, there was a steady increase in the number of surveys, with a significant rise in the frequency of income-related surveys. By the 2000s, the number of surveys received had peaked, with more than 40 surveys arriving annually in certain years. Throughout this period, income surveys consistently outnumbered consumption surveys. Nevertheless, the receipt of consumption data also increased. In recent years, there has been a notable decline in the number of surveys received, particularly post-2020 due to the impacts of the COVID-19 pandemic and longer processing and clearance times in the NSOs.

Figure 2: Frequency of household survey by type of welfare aggregate used



Note: The chart shows the number of surveys in the Global Monitoring Database for each calendar year. The surveys are disaggregated by the type of welfare aggregate – income or consumption – used. Surveys conducted in a fiscal year are reported as a survey in the starting calendar year. EUSILC surveys collected in year T are reported in year T-1 as income reference period refers to the previous calendar year.

Across 1,826 surveys, negative welfare was observed in 457 surveys—on average, 0.05 % of the population have negative welfare, with a maximum share of 1.30 %. Zero welfare occurred in 778 surveys—on average, 0.16 % of the population have zero welfare, with a maximum share of 10.68 %.

Table 2: Zero and negative welfare

	Share of population			Number of survey
	mean	min	max	
negative	0.05%	0.00%	1.30%	457
zero	0.16%	0.00%	10.68%	778

Note: The table reports the share of population and number of surveys influenced by negative or zero values.

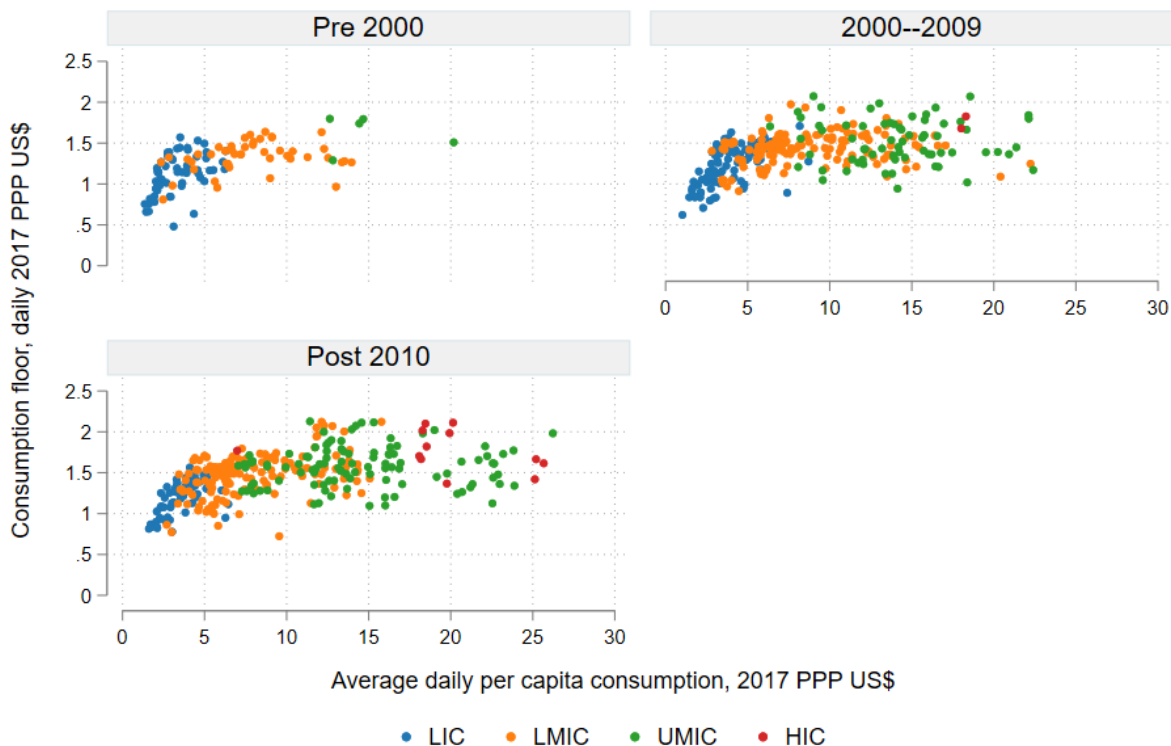
In what follows, unless otherwise mentioned, data is presented in daily household per capita 2017 PPP\$.

## 4. Results

### 4.1. Permanent consumption floor using the Ravallion (2016) method

The permanent consumption floor is a function of the poverty line ( $z$ ) and the degree of curvature or weights ( $\alpha$ ). Given the two parameters, a permanent consumption floor can be estimated for each survey. Figure 3 reports the distribution of such floors estimated for all consumption surveys using all observations below  $z = \text{International Poverty Line (IPL)}$  – i.e., \$2 .15 per person per day in 2017 PPPs – and with  $\alpha = 1$  (i.e., equally weighted average). Table 2 reports the summary statistics of floors using other combinations of the two parameters. The estimated floors in Figure 3 range from \$0.48 per capita per day to \$2.13 per capita per day. The minimum value of the floor has only increased very slightly over the three decades. The consumption floor on average increases initially with the increase in the mean income of the survey; however, the floor on average is relatively flat after roughly \$5 of mean daily per capita income. The average value of the floor among all consumption surveys is \$1.40 and the median is \$1.41.

Figure 3: Distribution of permanent consumption floors,  $\alpha = 1$  and  $z = IPL$



Note: The figure reports the distribution of permanent consumption floors for each consumption survey ( $N = 809$ ) defined using a poverty threshold ( $z$ ) equivalent to the international poverty line (*IPL*) – i.e., \$2.15 per person per day in 2017 PPPs – and a curvature parameter ( $\alpha$ ) equal to 1 (see equation 1 and Ravallion 2016). Surveys are grouped by the decade in which they were conducted. Consumption is reported in daily per capita 2017 PPP\$.

Acronyms: LIC – low-income country; LMIC – lower middle-income country; UMIC - upper middle-income country; HIC – high-income country.

Using the *IPL* as the cutoff threshold is somewhat arbitrary. Consider surveys where most of the population lives above the *IPL*. In those surveys, the floor would be defined by very few survey observations if any. The opposite would be true for surveys that have most of the population living below the *IPL*. While not having enough observations might be limiting, there is at the same time no reason to believe that using all the observations in the survey adds to the precision of the estimate of a floor.

To that end, various other thresholds are considered. A floor based on the national poverty line (*NPL*) might be preferred for a more nuanced treatment across all countries.<sup>7</sup> However, the *NPL* is limited due to its lack of definitional comparability across countries. For better comparability, we consider thresholds like the societal poverty line (*SPL*) or a threshold consumption of certain percentile of the consumption distribution. The latter will also more consistently capture an equal

<sup>7</sup> National poverty lines in low, lower-middle, and most upper-middle income countries are absolute poverty lines, while most high-income countries use a relative poverty threshold.

share of the population across all surveys to estimate the floor. Table 2 reports a range of estimates of the consumption floors using the national poverty line (*NPL*), and various other thresholds: *SPL*, consumption threshold of 10<sup>th</sup> percentile of earners (*q10*), and consumption threshold of 25<sup>th</sup> percentile of earners (*q25*). Other thresholds and floors using income surveys are considered in Table D1.

Furthermore, the curvature of the fit could be non-linear, i.e.,  $\alpha > 1$ . A floor can be estimated using higher values of  $\alpha$ , which gives more importance to the smaller values of consumption compared to larger values. Table 3 also reports estimates of the floor with two choices for  $\alpha$ . A priori, it is unclear what the optimal value of  $\alpha$  should be. We restrict the range of options to  $\alpha = \{1,4\}$ . Any value of  $\alpha$  within these bounds will generate a floor that is within the range of floors reported in Table 3.

The minimum value of the permanent consumption floor ranges from \$0.02 to \$0.48 when considering all surveys for  $\alpha$  between 1 and 4 and between \$0.10 and \$0.63 when using more recent (post 2000) sample of surveys with the same range for  $\alpha$ . In general, the floors estimated using higher poverty lines are larger compared to floors estimated with lower poverty lines. For example, the median floor with  $\alpha = 1$  and using the *SPL* as the poverty threshold is \$2.58 compared to the median floor of \$1.99 with  $\alpha = 1$  and using the consumption of the 10<sup>th</sup> percentile as the poverty threshold. The typical *SPL* threshold is \$4.42 while a typical *q10* threshold is \$3.34. This is mechanical as the floors with the Ravallion (2016) approach are a weighted average of consumption values; the floor is larger if larger consumption values are incorporated in this average. As expected, floors using  $\alpha = 4$  are smaller than floors using  $\alpha = 1$  as the former gives more weight to smaller values of consumption while the latter gives equal weight across all observations below the poverty threshold.

It is probable that smaller values of consumption are more important to estimate the floor of consumption. This is why the above formulation gives no weight to consumption values above the poverty threshold and likewise increases  $\alpha$ . In a similar vein, floors estimated on poorer countries might be a less biased estimate of a universal floor. To that end, Table 3 also reports floors using recent (post 2000) surveys only from low-income countries (LICs). For this subsample, the minimum value of the permanent consumption floor ranges between \$0.10 with  $z = q10$  and  $\alpha = 4$ , and \$0.63 when  $z = IPL$  or *SPL* and  $\alpha = 1$ .

While there is a wide range of possible values for the permanent consumption floors, it is important to note that the higher end of reasonable values is roughly between \$0.50 and \$0.60 per day. It would be reasonable to argue that a bottom code above this threshold would be too limiting on the survey.

Table 3: Permanent consumption floor using various poverty thresholds and curvature

z	$\alpha$	All surveys (N = 809)			Post-2000 surveys (N = 698)			LIC surveys (N = 165)		
		Min	p50	Mean	Min	p50	Mean	Min	p50	Mean
IPL	1	0.48	1.41	1.40	0.63	1.44	1.43	0.63	1.26	1.21
IPL	4	0.24	1.03	1.06	0.38	1.06	1.10	0.41	0.85	0.86
NPL	1	0.45	2.42	2.69	0.45	2.61	2.82	0.45	1.24	1.40
NPL	4	0.31	1.73	1.94	0.31	1.86	2.03	0.31	0.90	1.00
SPL	1	0.48	2.58	2.95	0.63	2.86	3.10	0.63	1.41	1.44
SPL	4	0.24	1.86	2.11	0.42	2.04	2.23	0.42	0.99	1.03
q10	1	0.04	1.99	2.36	0.16	2.27	2.52	0.16	0.85	0.89
q10	4	0.02	1.42	1.69	0.10	1.60	1.82	0.10	0.58	0.63
q25	1	0.13	2.56	3.08	0.25	2.96	3.28	0.25	1.12	1.16
q25	4	0.06	1.88	2.20	0.17	2.10	2.35	0.17	0.76	0.83

Note: This table reports the minimum, median, and mean value of consumption floors across surveys using the method proposed by Ravallion (2016) (see equation 1). The floors depend on a poverty threshold ( $z$ ) and a curvature parameter ( $\alpha$ ) used for weighting. A linear weight is assumed when  $\alpha = 1$ , a quadratic weight with higher weight to the bottom values when  $\alpha = 2$ , and so on. The summary statistics from the distribution of floors in Figure 1 are reported in the first 3 columns of the first row ( $z = \text{IPL}$  and  $\alpha = 1$ ). Floors are reported in daily per capita 2017 PPP\$. For floors using other thresholds ( $z$ ), curvature parameter ( $\alpha$ ), and income surveys, see also Table D1.

Acronyms:  $z$  - poverty line; Min - minimum; p50 - median; IPL - International poverty line; NPL - national poverty line; SPL - societal poverty line; q10 - threshold income of the 10<sup>th</sup> percentile; q25 - threshold income of the 25<sup>th</sup> percentile.

## 4.2. Consumption floor using various parametric distributions

### 4.2.1. Consumption floor using the Beta distribution

It is unclear which value of the threshold and  $\alpha$  is preferred in the Ravallion framework. The appeal for using the Beta distribution instead is the known functional form, which allows for only varying the threshold parameter. However, unlike floors estimated using Ravallion's approach, floors defined using the Beta distribution are not bounded at the bottom. Recall that Ravallion's floor is a weighted average of the consumption values of the bottom of the distribution; they are bounded at the bottom by design at the observed minimum value in the survey. The observed minimum is the extreme value of the floor when there is only one observation below the poverty threshold, or all the weight is assigned to this observation. This is not the case for Beta distribution. Indeed, depending on the poverty threshold used, between 1% and 10% of floors estimated with the Beta distribution approach are below zero. Table 4 presents summary statistics of the distribution of consumption floors excluding non-positive values. The typical consumption floor ranges from \$0.69 to \$1.12 when using all consumption surveys, from \$0.71 to \$1.29 when using more recent (post-2000) surveys, and from \$0.32 to \$0.76 when restricting the sample to recent LIC surveys. The median value of the consumption floor is robust to the inclusion of negative values (see Table D2).

Table 4: Consumption floors using the Beta distribution

z	All surveys				Post-2000 surveys				LIC surveys			
	N	Min	p50	Mean	N	Min	p50	Mean	N	Min	p50	Mean
IPL	693	0.02	0.69	2.13	594	0.02	0.71	1.93	160	0.05	0.62	2.37
NPL	783	0.01	1.09	1.22	673	0.01	1.18	1.27	164	0.14	0.65	0.75
SPL	795	0.01	1.10	1.27	686	0.01	1.19	1.33	162	0.09	0.76	0.77
q10	721	<0.01	0.73	0.95	633	<0.01	0.80	1.00	143	<0.01	0.32	0.36
q25	797	<0.01	1.12	1.38	689	<0.01	1.29	1.48	162	<0.01	0.48	0.53

Note: This table reports the values of the consumption floor using the Beta distribution as outlined in equation 2. Negative floors are excluded. Floors are reported in daily per capita 2017 PPP\$. For floors using other thresholds (z) and income surveys, see also Table D2.

Acronyms: z - poverty line; N - number of surveys; Min - minimum; p50 - median; IPL - International poverty line; NPL - national poverty line; SPL - societal poverty line; q10 - threshold income of the 10<sup>th</sup> percentile; q25 - threshold income of the 25<sup>th</sup> percentile.

#### 4.2.2. Floors using the three-parameter log normal and the Dagum III distributions

Cohen and Whitten (1980) outline various ways to estimate the three-parameter log normal. Among those they outline, they recommend using a version of the method of moments estimator (ME) when the third moment is less than one and a version of the localized maximum likelihood estimator (LMLE) when the third moment is greater or equal to one. Table 5 reports values of the floor using both approaches. Note that there are only three surveys that have skewness that is less than 1. About a third of the floors using the LMLE and a tenth of the floors using ME are non-positive. These floors are excluded from Table 5.

The median consumption floor using the LMLE ranges from \$0.50 for low-income countries to \$0.61 when considering all surveys or post-2000 surveys. The minimum value of the floor is much smaller, \$0.02 when considering all surveys and \$0.24 when considering just the LIC surveys. We were unable to fit the maximum likelihood estimator to all distributions, hence, the floors are only reported for 34 LIC surveys. While we were able to apply the ME approach to a larger set of surveys, the results here must be considered in light of the recommendation from Cohen and Whitten (1980). The range of floors using the ME is generally higher than the floors using the LMLE. The median consumption floor using the ME ranges from \$0.87 for low-income countries to \$2.52 for all surveys after 2000. The minimum values of the floor are closer to zero. Note that the table also reports values fitted on income distribution as the TPLN and the Dagum III distributions have typically been used to fit income distributions in the past. The minimum value of the income floor is \$0.08 using the LMLE approach, while the median ranges between \$0.45 and \$0.51 depending on the time frame of the surveys. The floors are much larger using the ME for income surveys (the median floor ranging between \$1.20 and \$7.90), although there are only 8 income surveys in our LIC sample.

Table 5: Floors using the three-parameter log normal distribution

Sample	TPLN, LMLE			TPLN, ME				
	N	Min	p50	Mean	N	Min	p50	Mean
(a) Consumption surveys								
All	350	0.02	0.61	0.69	750	<0.01	2.28	2.82
Post-2000	327	0.02	0.61	0.70	644	<0.01	2.52	3.01
LIC	34	0.24	0.50	0.52	207	<0.01	0.87	1.07
(b) Income surveys								
All	34	0.08	0.45	0.64	870	0.01	7.10	12.41
Post-2000	31	0.08	0.51	0.69	798	0.01	7.90	13.33
LIC	-	-	-	-	8	0.28	1.20	1.13

Note: This table reports values of the floor for the three-parameter log normal (TPLN) distribution using the local maximum likelihood estimator (LMLE) and the method of moments estimates (MME). Negative floors are excluded. Floors are reported in daily per capita 2017 PPP\$.  
 Acronyms: LIC – low-income country; N – number of surveys; Min – minimum; p50 – median.

Like the TPLN, not all surveys could be fit using the maximum likelihood approach of the Dagum III distribution. Among those that did fit, a third of the floors were negative. Since Dagum III should strictly have positive floors, Table 6 presents results of the Dagum III for only positive floors. The median value of the floors estimated on consumption distributions ranges from \$0.22 for LIC surveys to \$0.79 for all surveys after 2000. The median of the floor for income distribution is \$0.58. For both consumption and income distributions, the minimum values of the positive floor are close to zero.

Table 6: Floors using the Dagum III distribution

	N	Min	p50	Mean
(a) Consumption surveys				
All	312	0.01	0.73	0.91
Post-2000	280	0.01	0.79	0.95
LIC	44	0.01	0.22	0.29
(b) Income surveys				
All	31	<0.01	0.58	0.71
Post-2000	31	<0.01	0.58	0.71

Note: This table reports values of the floor for the Dagum III distribution (see Appendix C). Negative floors are excluded. Floors are reported in daily per capita 2017 PPP\$.

Acronyms: LIC – low-income country; N – number of surveys; Min – minimum; p50 – median.

#### 4.3. Floor using outlier detection method

Floors can be defined using outlier detection methods a la LIS approach defined in Section 3.3. Table 7 reports consumption and income floors defined at  $k$  times the  $IQR$  distance below the

first quartile of the distribution (see equation 3). LIS uses  $k = 3$ . Table 7 reports values for  $k$  between 2 and 4. The minimum value of the floors using this approach is quite small, in all cases less than \$0.05. The median among consumption surveys ranges from \$0.16 (when  $k = 4$ ) to \$0.85 (when  $k = 2$ ) using all surveys, from \$0.20 (when  $k = 4$ ) to \$0.96 (when  $k = 2$ ) when considering surveys that were conducted after 2000, and from \$0.06 (when  $k = 4$ ) to \$0.32 (when  $k = 2$ ) when focusing only on low-income countries. The income floors are lower. This is not unexpected as in any distance-based approach, like using the *IQR*, the floors are lower if the *IQR* is higher. Inequality, in this case *IQR*, in income distributions are typically higher than consumption distributions. The typical income floor ranges on the low end from \$0.01 (when  $k = 4$ ) to \$0.05 (when  $k = 2$ ) when using only LIC surveys or from \$0.58 (when  $k = 4$ ) to \$2.91 (when  $k = 2$ ) when considering all country surveys after 2000.

Table 7: Floors using LIS-type method

$k$	All surveys				Post-2000 surveys				LIC surveys			
	N	Min	p50	Mean	N	Min	p50	Mean	N	Min	p50	Mean
(a) Consumption surveys												
2	809	0.01	0.85	1.10	698	0.05	0.96	1.18	165	0.05	0.34	0.38
3	809	<0.01	0.36	0.53	698	0.01	0.44	0.57	165	0.01	0.14	0.18
4	809	<0.01	0.16	0.26	698	0.00	0.20	0.28	165	<0.01	0.06	0.08
(b) Income surveys												
2	1017	0.03	2.01	3.89	905	0.03	2.91	4.32	3	0.04	0.05	0.12
3	1017	0.01	0.84	1.96	905	0.01	1.32	2.18	3	0.01	0.01	0.04
4	1017	<0.01	0.35	1.00	905	0.00	0.58	1.12	3	<0.01	0.01	0.01

Note: This table reports statistics of the floor using the expression:  $y_f = \exp(Q_1 - k \times (Q_3 - Q_1))$ , where  $Q_1$  is the 1<sup>st</sup> quartile,  $Q_3$  is the 3<sup>rd</sup> quartile, and  $k$  is the distance parameter. While the table reports the floors using  $k \in \{2,3,4\}$ , LIS uses  $k = 3$ . Floors are reported in daily per capita 2017 PPP\$.

Acronyms: N – number of surveys; Min – minimum; p50 - median.

#### 4.4. Minimum cost of least expensive energy sufficient diet

Beyond statistical approaches, the food poverty literature could also shed some light into the biological consumption floor. This literature generally classifies the cost of diet into three categories: energy sufficient diet, nutrient adequate diet, and healthy diet. Energy sufficient diet is the minimum cost to meet energy requirements using the least-cost available, and nutrient adequate and healthy diet supplements the former. Energy sufficient diet are typically starchy staple food items in each country. The FAO defines energy cost diet as the intake that “meets the need for short-term subsistence” (FAO 2020). This category could be used to understand the average minimum requirement across countries. Figure 4 reports the energy sufficient diet cost across countries. The costs range from \$0.24 per day to close to \$3.00 per day. Surprisingly the costs of this diet is fairly stable across countries at various income levels with the average cost in the richest countries almost equal to the average cost in the poorest set of countries. The median cost among low-income countries is \$0.76 per day. If instead the focus was on the bottom 10% of

the minimum energy costs (note this would include countries in the high-income category as well), the median floor is \$0.31 per day.

Figure 4: Distribution of the average energy sufficient diet costs of countries by income level



Note: This figure reports the distribution of daily costs of an energy sufficient diet using 2017 PPP\$. Costs and income are reported in daily per capita 2017 PPP\$.

## 5. Discussion

### 5.1. The threshold to bottom code

The bottom code must consider two constraints: on the one hand, a floor that is too large implies that a large segment of the distribution is affected by the bottom code. Table D3 reports the share of population of income and consumption surveys affected by a \$0.25 and \$0.50 per day bottom code for the top 10 affected countries. The higher the bottom code the larger the share of data impacted by the bottom code. For example, a bottom code of \$0.50 per day would mean altering up to 30% of the data in the Democratic Republic of Congo’s 2004 survey, whereas a bottom code of \$0.25 per day would affect “only” 6% of the data. This would suggest that erring on the conservative side of bottom code is preferable.

On the other hand, a floor that is too small implies that a small part of the distribution can become very influential for (bottom-sensitive) measures. Additionally, small observations are likely to be prone to measurement errors.

### 5.1.1. Evidence from the cost of energy sufficient diet and statistical approaches

Herforth et al. (2022) use prices from the International Comparison Program to estimate the cost of a diet consisting of the least expensive staple starches that results in 2,330 kcal per day. The minimum across all countries is \$0.24 per person per day in 2017 PPP\$. While someone can survive on fewer calories for some periods of time, it should also be noted that this diet is not healthy as it only covers calories without accounting for other nutrients. Interviews conducted on a handful of families over a week in December 2020 in Yemen revealed that people lived on 1,332 kcal per day during the conflict (Liskey 2021). This is roughly half of the recommended daily intake of 2,330 kcal. In the median low-income country, it would cost around \$0.38/day to consume 1,332 kcal per day (the median cost among LICs for 2,330 kcal is \$0.76/day).

With statistical approaches, the distribution of floors varies depending on the choice of method or underlying parameter. If we focus on the recent surveys in low-income countries with a poverty threshold at the 10<sup>th</sup> percentile (or 25<sup>th</sup> percentile), the typical Beta distribution floor was \$0.32 (\$0.48), while the smallest value of the permanent consumption floor was \$0.16 (\$0.25) considering equally distributed weights (Table 8). The typical consumption floor for LICs ranged between \$0.24 and \$0.87 when considering the three-parameter log normal and \$0.22 when considering the Dagum III distribution. The typical LIS-type floor for recent LIC surveys was \$0.14.

Table 8: Summary of the floors using various statistical approaches

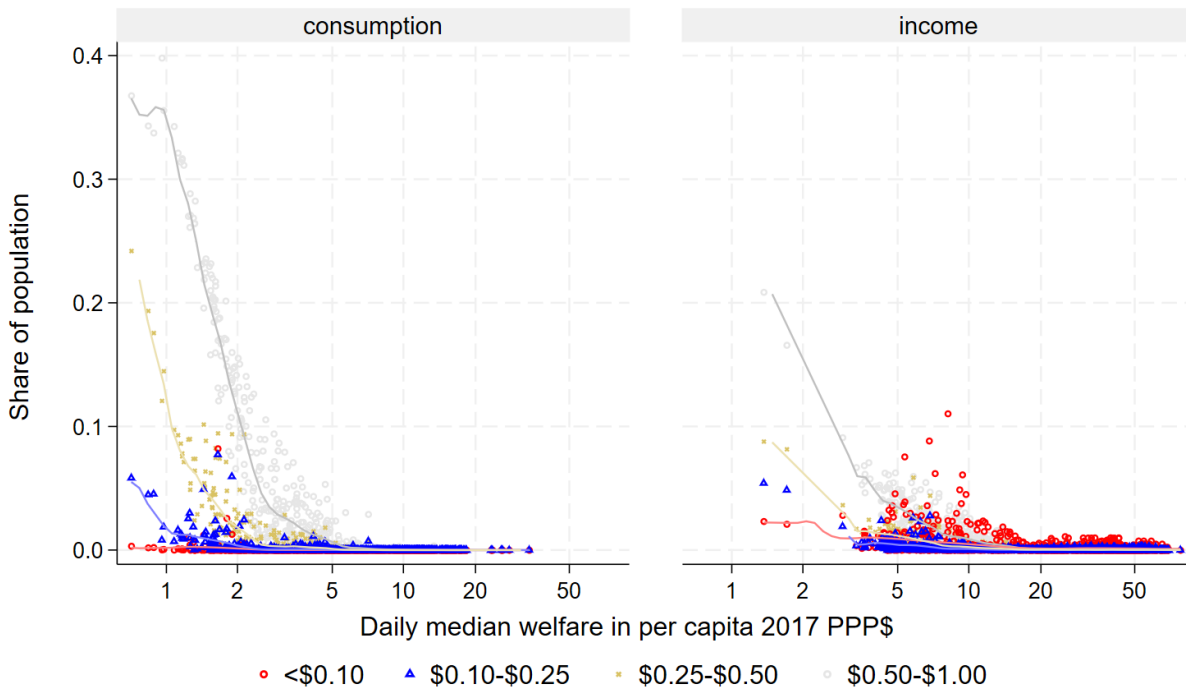
Method	Value of floor in daily 2017 PPPs		
	All data/surveys	Post-2000 surveys	LIC surveys
I. Ravallion ( $\alpha = [1,4]$ ; $z = q10$ )	\$ [0.02, 0.04]	\$ [0.10, 0.16]	\$ [0.10, 0.16]
	<i>N</i> 809	698	165
II. Beta ( $z = p10$ )	\$ 0.73	\$ 0.80	\$ 0.32
	<i>N</i> 721	633	143
III. TPLN	\$ 0.61	\$ 0.61	\$ 0.50
	<i>N</i> 350	327	34
IV. Dagum III	\$ 0.73	\$ 0.79	\$ 0.22
	<i>N</i> 312	280	44
V. LIS (3 x IQR below Q1)	\$ 0.36	\$ 0.44	\$ 0.14
	<i>N</i> 809	698	165

Note: This table summarizes consumption floors across various methods outlined in Section 4. Ravallion floor is the minimum value, while all other floors are median values. See Tables 2, 3, 4, 5, and 6. Floors are reported using daily per capita 2017 PPP\$. Acronyms: *z* - poverty line; *q10* – threshold income of the 10<sup>th</sup> percentile; *N* – sample size.

### 5.1.2. Evidence from inspecting the data

One could also inspect the raw data for measurement concerns. One crude approach is to correlate the share of population with small values of income or consumption to the median welfare in the survey. Barring measurement concerns, we should expect the share of people with small value of consumption and income to decrease as the median welfare of distribution rises. To see this, Figure 5, plots the distribution of consumption and income in each survey for those below \$0.10, between \$0.10 and \$0.25, between \$0.25 and \$0.50, and between \$0.50 and \$1.00 per day. The share of the population with those amounts of consumption or income is compared to the median of the survey. For the first two lower value groups, there are some noticeable jumps. For instance, the share of people earning less than \$0.10 increases in many surveys when the daily median income is between \$5 and \$10. Compared to this, the share having below \$0.10 is much less in surveys with median income less than \$5. Similar erratic jumps are evident for the next higher bin. The share of people consuming between \$0.10 and \$0.25 increases when the median consumption is around \$2; this proportion is much lower in surveys with median consumption around \$1. These patterns are not as visible once one moves to higher consumption and income bins, i.e., those above \$0.25 per day. The erratic jumps for the bins with lower welfare value could be potential signs of greater measurement issues.

Figure 5: Share of population in income groups below various thresholds



Note: This figure reports the share of people living with various amounts of consumption or income in each survey expressed in daily 2017 PPP\$.

### 5.1.3. Evidence from contribution function

Another approach is to check the contribution of some parts of the distribution on an indicator. Since the coding is at the bottom of the distribution, we can take a distribution sensitive indicator and check how much the bottom  $x\%$  of the distribution contributes to the value of the indicator with and without various bottom codes. We then compare this contribution to the contribution on the same indicator using the  $x\%$  of a well-known distribution.

In what follows, we estimate the contribution of the bottom decile to the Prosperity Gap (Kraay et al. 2023) for a well-behaved distribution (the lognormal) and compare this with the contribution of the bottom decile across our sample with various bottom codes. We can define the contribution of the  $p^{th}$  percentile to the Prosperity Gap as  $S(p) \equiv \frac{\sum_{i=1}^p y_i^{-1}}{\sum_{i=1}^N y_i^{-1}}$  with income,  $y$ , and  $i = 1, 2, \dots, N$  observations. For a lognormal distribution,  $S(p) = \Phi(\sigma - \Phi^{-1}(1 - p))$ , where  $p$  is the percentile of interest and standard deviation,  $\sigma = \sqrt{2}\Phi^{-1}\left(\frac{1+G}{2}\right)$ , and  $G$  is the Gini coefficient. This implies that the contribution of the bottom decile to the Prosperity Gap in a country with very high Gini coefficient (Gini index equal to 60) is 46%. Likewise, the bottom decile's contribution for a country with Gini of 40 is 30%. In our sample of 1,826 surveys, about 1% of surveys have a Gini index as high or higher than 60, 35% have Gini between 40 and 60, and 65% have Gini less than 40. So, given a lognormal assumption, it would be quite rare to find the bottom decile contributing much more than 46% while it is quite common to find contributions less than 30% (in 2/3<sup>rd</sup> cases) or between 30% and 46% (in almost 1/3<sup>rd</sup> cases).

Table 9 reports the contribution of the bottom decile to the Prosperity Gap for consumption and income surveys. The distribution of contribution is grouped into the share of surveys with less than 30%, between 30% and 45%, and higher than 45% contribution. In the raw data (panel a), we find that in 88% (26%) of consumption (income) surveys, the bottom decile contributes less than 30% to the Prosperity Gap. We also see that in 1% of consumption surveys and 22% of income surveys, the bottom decile contributes more than 45% to the Prosperity Gap.

Bottom coding at any level limits the number of surveys with contribution in excess of 45%. For instance, bottom coding at \$0.10 brings down the share of income surveys with contribution over 45% from 22% to 13%. If income surveys more closely follow a lognormal distribution, then it could be argued that the higher the bottom code, the more closely it mimics the distribution of Ginis in our sample. For instance, a bottom code of \$0.50 would imply that 35% of income surveys have contributions less than 30% to the Prosperity Gap, 62% of income surveys have contributions between 30% and 40%, and only 3% of income surveys have contributions above 45%.

Table 9: The distribution of contribution of the bottom decile to the Prosperity Gap

	Share of consumption surveys			Share of income surveys		
	below 30%	30% - 45%	above 45%	below 30%	30% - 45%	above 45%
(a) Raw data	88.4	10.5	1.1	25.8	52.5	21.8
(b) Threshold						
\$0.10	88.6	10.8	0.6	27.8	59.0	13.3
\$0.15	89.1	10.5	0.4	28.6	61.0	10.3
\$0.20	89.7	10.0	0.2	29.5	62.1	8.4
\$0.25	90.4	9.5	0.1	30.3	62.4	7.3
\$0.30	91.3	8.7	0.0	31.0	62.8	6.2
\$0.35	92.1	7.9	0.0	31.9	62.9	5.2
\$0.40	93.0	7.0	0.0	32.6	63.0	4.4
\$0.45	93.7	6.3	0.0	33.1	62.8	4.1
\$0.50	94.2	5.8	0.0	34.5	62.2	3.2

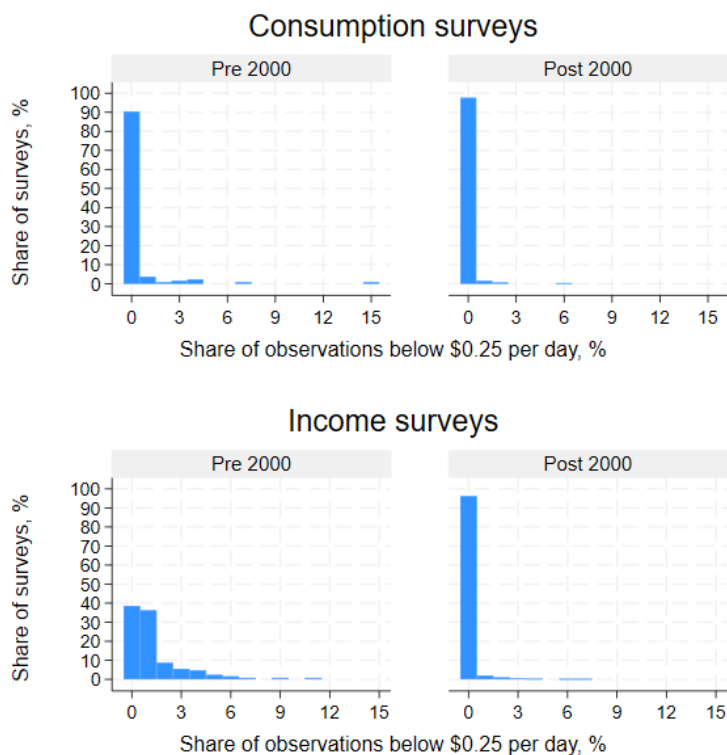
Note: This table reports the share of surveys with the bottom decile contributing less than 30%, between 30% and 45%, and over 45% to the Prosperity Gap in the survey.

## 5.2. Impact of a bottom code at \$0.25 a day

### *Share of population below various bottom codes and the differences in welfare measures*

Figure 6 shows the share of population below the \$0.25 per person per day bottom code in consumption and income surveys separated into pre and post 2000 surveys. While there are extreme cases, for example the Zambia 1991 survey has over 15% of the population below \$0.25 per day, this is an exception rather than a norm. In over 95% of post-2000 surveys and at least 90% of all surveys in the GMD, there are no observations below \$0.25. While there are more than half the pre-2000 income surveys that have some part of the distribution below \$0.25, only 14 surveys or 11% of pre-2000 income surveys have more than 3% of the population below the bottom code, with the maximum being 11% below the bottom code in Suriname's 1999 income survey.

Figure 6: Distribution of share-of-population-below \$0.25 per day across surveys



Note: This figure reports the share of surveys with observations below \$0.25 per day.

### *Sensitivity of poverty and inequality indicators to various thresholds*

Table 10 reports the difference in mean income, Gini index, mean log deviation (MLD), and the Prosperity Gap due to the bottom code for the ten surveys with the most observations below \$0.25 per day. Panel a reports the estimates for consumption surveys and panel b reports the same for income surveys. The difference in estimated values of the mean income is less than 0.5% in most cases with a difference in mean income of 0.7% in Zambia 1991 where the mean income before the bottom code is \$3.10 and after the bottom code is \$3.12. The differences in the Gini index are slightly larger but limited to consumption surveys. Again, the largest difference, 1.8%, is for Zambia 1991 where the Gini prior to the bottom code is 60.5 and after is 59.4.

As expected, the differences for the bottom sensitive measures are larger. For example, the Prosperity Gap for the Zambia 1991 survey changes from 212 before the bottom code to 33 after (84% lower) and the MLD changes from 0.937 before to 0.735 after (22% lower). However, the

correlation between the Prosperity Gap and the mean income is much higher after the adjustment compared to before, implying that the rankings are more in line with expectations and not influenced by a few very small observations.

## 6. Conclusion

Triangulating various evidence and with the constraints (too high or too low bottom code) in mind, a threshold of \$0.25 per day seems reasonable to use as a bottom code from the range of methods discussed above. A threshold above \$0.50 per day would be too intrusive on the surveys considering that the higher end of permanent consumption ranges around this threshold. The range of values between \$0.25 and \$0.50 limits observable measurement errors, while a threshold below \$0.25 per day does not fully limit the contribution of bottom on distribution sensitive indicators compared to a well-behaved distribution. The \$0.25 per day threshold in 2017 PPPs is also almost equivalent to the minimum of the least-cost energy sufficient diet, and it is the range of the typical floors estimated using various statistical approaches.

Given the wide range of estimates of the floor with various methodologies, a bottom code defined on each survey could be preferable. However, even for the same country, the range of estimated floors can be quite different depending on the survey if one were to use such an approach. More importantly, bottom codes that vary across countries or within a country across time imply a range of possible outcomes for most poverty and inequality measures. For example, the Prosperity Gap with a bottom code at \$0.25 per day has a maximum factor of 100 while the Prosperity Gap with a bottom code at \$0.50 per day will have a maximum factor of 200.<sup>8</sup> Comparing estimates defined on indicators that have different domains is not ideal.

---

<sup>8</sup> The Prosperity Gap estimated using the \$25 a day target threshold. Likewise, the Watts Index has a domain between 0 and 2.15 or 0 and 1.46 depending on the choice of \$0.25 or \$0.50 to bottom code with a poverty threshold at \$2.15.

Table 10: Estimates for various indicators with and without bottom code at \$0.25 per day

Country	Year	Pop below \$0.25, %	Mean	Mean*	Diff, %	Gini	Gini*	Diff, %	MLD	MLD*	Diff, %	PG	PG*	Diff, %
<b>(a) Consumption surveys</b>														
Zambia	1991	15.9	3.10	3.12	0.7%	60.5	59.4	-1.8%	0.937	0.735	-21.6%	212.0	33.0	-84%
Lesotho	1994	7.1	4.33	4.34	0.1%	63.2	63.0	-0.3%	0.839	0.804	-4.1%	49.1	25.5	-48%
Lesotho	2002	6.1	2.27	2.27	0.2%	51.6	51.2	-0.7%	0.511	0.480	-6.1%	33.3	27.5	-17%
Dem. Rep. Congo	2004	6.1	1.00	1.01	0.4%	42.2	41.6	-1.2%	0.301	0.285	-5.4%	44.1	41.2	-6%
Mozambique	1996	4.8	1.59	1.59	0.2%	53.6	53.3	-0.5%	0.498	0.485	-2.6%	37.0	35.0	-5%
Mali	1994	4.7	1.42	1.43	0.2%	50.4	50.2	-0.6%	0.437	0.424	-2.8%	37.8	35.7	-5%
Uzbekistan	1998	4.3	2.45	2.45	0.1%	43.7	43.6	-0.4%	0.349	0.327	-6.3%	29.7	20.3	-31%
The Gambia	1998	3.6	1.93	1.93	0.2%	48.5	48.3	-0.6%	0.422	0.402	-4.8%	30.7	26.8	-13%
Burundi	1998	3.4	1.65	1.66	0.2%	42.3	42.1	-0.6%	0.326	0.309	-5.1%	31.1	27.1	-13%
Benin	2015	3.0	3.00	3.01	0.1%	47.8	47.6	-0.3%	0.446	0.431	-3.3%	23.0	20.4	-12%
<b>(b) Income surveys</b>														
Suriname	1999	11.0	13.1	13.1	0.0%	52.8	52.8	0.0%	0.430	0.418	-2.9%	14.5	5.37	-63%
Panama	1989	9.8	11.9	11.9	0.0%	55.0	55.0	0.0%	0.540	0.533	-1.2%	9.53	8.27	-13%
Colombia	2000	8.0	10.0	10.0	0.0%	55.3	55.3	0.0%	0.481	0.478	-0.6%	7.23	6.69	-7%
Haiti	2001	7.7	2.67	2.68	0.3%	59.5	59.0	-0.8%	0.692	0.640	-7.6%	41.5	28.6	-31%
Panama	1991	7.0	12.7	12.7	0.0%	55.5	55.5	0.0%	0.581	0.576	-1.0%	9.03	8.09	-10%
Haiti	2012	7.0	3.49	3.50	0.2%	60.6	60.3	-0.5%	0.732	0.688	-6.0%	39.6	24.9	-37%
Panama	1995	6.7	16.2	16.2	0.0%	55.1	55.1	0.0%	0.553	0.550	-0.6%	6.68	6.05	-9%
Micronesia	2000	5.8	5.60	5.60	0.1%	63.3	63.1	-0.3%	0.857	0.817	-4.7%	31.3	21.1	-33%
Panama	1998	5.3	16.1	16.1	0.0%	55.4	55.4	0.0%	0.566	0.562	-0.6%	6.59	5.99	-9%
Colombia	1999	5.2	8.98	8.98	0.0%	56.8	56.8	0.0%	0.561	0.558	-0.5%	9.11	8.71	-4%

Note: This table reports difference (in percent) in the mean, Gini index, mean log deviation (MLD), and Prosperity Gap (PG) for indicators calculated without a bottom code and those calculated with a bottom code at \$0.25 (\*). Negative observations are not included. Panel (a) reports the top 10 consumption surveys affected by the bottom code and panel (b) reports the same for income surveys. *Acronym*: Diff – difference.

## Appendix A: Beta distribution floor

Assume that consumption levels between the consumption floor and poverty line,  $z$ , follow a shifted and rescaled Beta distribution, i.e.,  $y = y_f + (z - y_f)x$  with  $x \sim \text{Beta}(\eta, \beta)$ . As long as  $z$  is below the modal value of the overall density of the distribution, it is reasonable to impose the further restrictions that  $\eta > 1$  and  $\beta = 1$ . Together, this ensures that the density of  $y$  is monotonically increasing over the interval  $[y_f, z]$ . The moment conditions for the Beta distribution with  $\eta > 1$  and  $\beta = 1$  are as follows:

$$E[y] = y_f + (z - y_f) \frac{\eta}{\eta + 1}$$

$$\text{var}[y] = (z - y_f)^2 \frac{\alpha}{(\eta + 1)^2(\eta + 2)}.$$

Poverty gap,  $P_1$ , is

$$P_1 = \frac{1}{N} \sum_i^q \left(1 - \frac{y_i}{z}\right)$$

where,  $N$  is total number of individuals  $i$ ,  $Q \in \{1, 2, \dots, q\}$  is the total number of poor, and  $z$  is poverty line.  $P_1$  can be rewritten as

$$P_1 = \left(\frac{1}{N} \sum_i^q 1\right) - \left(\frac{1}{z} \times \frac{1}{N} \sum_i^q y_i\right)$$

$$P_1 = H - \left(\frac{1}{z} \times \frac{Q}{N} \times \frac{1}{Q} \sum_i^q y_i\right) = P_0 - \frac{P_0}{z} E[y]$$

$$P_1 = P_0 \left(1 - \frac{E[y]}{z}\right). \quad (i)$$

Note that  $E[y]$  is the mean income of those who are poor and  $P_0$  is the headcount ratio of the poor. The ratio of poverty gap to headcount then is  $\frac{P_1}{P_0} = \left(1 - \frac{E[y]}{z}\right)$ . Using this ratio in equation (i), we can rewrite the expectation as  $E[y] = z \left(1 - \frac{P_1}{P_0}\right)$ .

Likewise, the squared poverty gap,  $P_2$ , can be written as

$$P_2 = \left(\frac{1}{N} \sum_i^q \left(1 - \frac{y_i}{z}\right)^2\right) = \frac{1}{N} \sum_i^q 1 - \frac{2}{z} \frac{1}{N} \sum_i^q y_i + \left(\frac{1}{z^2} \times \frac{1}{N} \sum_i^q y_i^2\right)$$

$$P_2 = P_0 - \frac{2E[y]P_0}{z} + \frac{E[y^2]P_0}{z^2}.$$

Hence the ratio of  $P_2$  to headcount can be rewritten as

$$\frac{P_2}{P_0} = 1 - \frac{2E[y]}{z} + \frac{E[y^2]}{z^2}$$

$$\frac{P_2}{P_0} - 1 = -\frac{2E[y]}{z} + \frac{E[y^2]}{z^2}$$

and

$$E[y^2] = \frac{P_2 z^2}{P_0} - z^2 + 2E[y]z.$$

The variance then can be written as

$$\begin{aligned} \text{var}(y) &= E[y^2] - E[y]^2 = \frac{P_2 z^2}{P_0} - z^2 + 2E[y]z - z^2 + \frac{2z^2 P_1}{P_0} - \frac{z^2 P_1^2}{P_0^2} \\ \text{var}(y) &= \frac{P_2 z^2}{P_0} - z^2 + 2\left(z - \frac{z P_1}{P_0}\right)z - z^2 + \frac{2z^2 P_1}{P_0} - \frac{z^2 P_1^2}{P_0^2} \\ \text{var}(y) &= \frac{P_2 z^2}{P_0} - z^2 + 2z^2 - 2\frac{z^2 P_1}{P_0} - z^2 + \frac{2z^2 P_1}{P_0} - \frac{z^2 P_1^2}{P_0^2} \\ \text{var}(y) &= \frac{P_2 z^2}{P_0} - \frac{z^2 P_1^2}{P_0^2}. \end{aligned}$$

Solving for floor using the Beta distribution using the moment conditions,

$$\begin{aligned} E[y] &= y_f + \frac{(z - y_f)\eta}{\eta + 1} \\ z\left(1 - \frac{P_1}{P_0}\right) &= \frac{\eta y_f + y_f + \eta z - \eta y_f}{\eta + 1} \\ (P_0 z - P_1 z)(\eta + 1) &= (y_f + \eta z)P_0 \\ \eta P_0 z + P_0 z - \eta P_1 z - P_1 z &= y_f P_0 + \eta z P_0 \\ P_0 z - \alpha P_1 z - P_1 z &= y_f P_0 \\ y_f &= z - \frac{\eta P_1 z}{P_0} - \frac{P_1 z}{P_0} = z\left(1 - (1 + \eta)\frac{P_1}{P_0}\right) \end{aligned}$$

$$\text{Or, } (z - y_f) = \frac{(\eta + 1)P_1 z}{P_0}$$

Using variance from the Beta distribution,

$$\text{var}(y) = \frac{(z - y_f)^2 \eta}{(\eta + 1)^2 (\eta + 2)}$$

$$\frac{P_2 z^2}{P_0} - \frac{z^2 P_1^2}{P_0^2} = \frac{(\eta + 1)^2 P_1^2 z^2 \eta}{(\eta + 1)^2 (\eta + 2) P_0^2}$$

$$(P_0 P_2 z^2 - z^2 P_1^2)(\eta + 2) = P_1^2 z^2 \eta$$

$$\eta P_0 P_2 z^2 + 2 P_0 P_2 z^2 - \eta z^2 P_1^2 - 2 z^2 P_1^2 = \eta z^2 P_1^2$$

Which gives the value for alpha,

$$\eta = \frac{2z^2 P_1^2 - 2z^2 P_0 P_2}{P_0 z^2 P_2 - 2z^2 P_1^2} = \frac{2P_1^2 - 2P_0 P_2}{P_0 P_2 - 2P_1^2}$$

Solving for floor,

$$y_f = z - \frac{\eta P_1 z}{P_0} - \frac{P_1 z}{P_0}$$

$$y_f = z - \left( \frac{2P_1^2 - 2P_0 P_2}{P_0 P_2 - 2P_1^2} \right) \frac{P_1 z}{P_0} - \frac{P_1 z}{P_0}$$

$$= z - \left( 1 + \frac{2P_1^2 - 2P_0 P_2}{P_0 P_2 - 2P_1^2} \right) \frac{P_1 z}{P_0}$$

$$= z - \left( \frac{P_0 P_2 - 2P_1^2 + 2P_1^2 - 2P_0 P_2}{P_0 P_2 - 2P_1^2} \right) \frac{P_1 z}{P_0}$$

$$= z - \frac{-P_0 P_2}{P_0 P_2 - 2P_1^2} \frac{P_1 z}{P_0}$$

$$y_f = z \left( 1 + \frac{P_0 P_1 P_2}{P_0^2 P_2 - 2P_0 P_1^2} \right).$$

## Appendix B: Three-parameter log normal

Following Cohen & Whitten (1980) the likelihood function can be expressed as

$$L(x; \gamma, \mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \cdot \prod_1^n (x_i - \gamma)^{-1} \cdot \exp \left\{ -\frac{[\ln(x_i - \gamma) - \mu]^2}{2\sigma^2} \right\},$$

where,  $\sigma > 0$ ,  $-\infty < \mu < \infty$ , and  $\gamma < x < \infty$ .  $\mu$  is the mean,  $\sigma^2$  is the variance, and  $\gamma$  is a threshold parameter. Then the function to maximize is

$$\ln f = \ln(x_i - \gamma).$$

## Appendix C: Dagum III distribution

Following Kleiber (2007), the probability density function can be expressed as

$$f(x) = \frac{apx^{ap-1}}{b^{ap} \left[1 + \left(\frac{x}{b}\right)^a\right]^{p+1}},$$

where,  $a, b, p > 0$  and  $0 < x_0 \leq x < \infty$ . The maximum likelihood function can be expressed as follows:

$$\ln f = \ln a + \ln p - ap \ln b + (ap - 1) \ln x - (p + 1) \ln \left[1 + \left(\frac{x}{b}\right)^a\right].$$

## Appendix D: Other results

Table D1: Permanent floors using Ravallion (2016)

z	$\alpha$	All surveys				Post-2000 surveys				LIC surveys			
		N	Min	p50	Mean	N	Min	p50	Mean	N	Min	p50	Mean
(a) Consumption surveys													
q05	1	809	0.01	1.69	1.99	698	0.11	1.93	2.13	165	0.11	0.70	0.75
q05	4	809	0.01	1.20	1.44	698	0.07	1.34	1.55	165	0.07	0.48	0.53
q50	1	809	0.36	3.31	4.01	698	0.36	3.73	4.27	165	0.36	1.50	1.53
q50	4	809	0.19	2.44	2.88	698	0.24	2.74	3.07	165	0.24	1.04	1.10
(b) Income surveys													
IPL	1	591	0.00	0.92	0.91	479	0.00	0.95	0.95	3	0.69	0.70	0.82
IPL	4	591	0.00	0.40	0.44	479	0.00	0.43	0.47	3	0.38	0.39	0.47
NPL	1	729	1.25	4.03	6.05	617	1.54	4.21	6.28	3	-	-	-
NPL	4	729	0.26	2.16	2.99	617	0.26	2.24	3.10	3	-	-	-
SPL	1	714	0.69	3.44	4.96	602	0.69	3.96	5.49	3	0.69	0.70	1.00
SPL	4	714	0.03	1.81	2.41	602	0.03	1.99	2.66	3	0.38	0.39	0.58
q05	1	656	0.00	1.39	2.78	544	0.02	1.69	3.22	3	0.06	0.08	0.24
q05	4	656	0.00	0.66	1.22	544	0.00	0.81	1.41	3	0.02	0.05	0.14
q10	1	693	0.01	2.03	3.92	581	0.05	2.50	4.49	3	0.13	0.13	0.36
q10	4	693	0.00	1.02	1.83	581	0.01	1.27	2.10	3	0.06	0.07	0.20
q25	1	746	0.26	3.69	6.30	634	0.26	4.36	7.09	3	0.26	0.29	0.62
q25	4	746	0.07	1.94	3.21	634	0.07	2.30	3.62	3	0.14	0.15	0.35
q50	1	828	0.49	6.55	9.92	716	0.49	7.73	10.98	3	0.49	0.57	1.02
q50	4	828	0.18	3.57	5.35	716	0.18	4.06	5.92	3	0.26	0.31	0.60

Note: This table reports the minimum, median, and mean value of consumption floors across surveys using the method proposed by Ravallion (2016) (see equation 1). The floors depend on a poverty threshold ( $z$ ) and a curvature parameter ( $\alpha$ ) used for weighting. A linear weight is assumed when  $\alpha = 1$ , a quadratic weight with higher weight to the bottom values when  $\alpha = 2$ , and so on. Floors are reported in daily per capita 2017 PPP\$. Negative income observations are not included. See also Table 3.

Acronyms: z - poverty line; Min - minimum; p50 - median; IPL - International poverty line; NPL - national poverty line; SPL - societal poverty line; q05 - threshold income of 5<sup>th</sup> percentile; q10 - threshold income of the 10<sup>th</sup> percentile; q25 - threshold income of the 25<sup>th</sup> percentile; q50 - threshold income of 50<sup>th</sup> percentile.

Table D2: Floors using Beta distribution

z	All surveys				Post-2000 surveys				LIC surveys			
	N	Min	p50	Mean	N	Min	p50	Mean	N	Min	p50	Mean
(a) Consumption surveys												
IPL	809	-202.3	0.60	0.70	698	-202.3	0.62	0.39	165	-2.23	0.61	2.27
NPL	809	-73.29	1.05	0.96	698	-73.29	1.12	0.98	165	-0.25	0.65	0.75
SPL	809	-17.85	1.09	1.21	698	-17.85	1.16	1.27	165	-1.14	0.75	0.75
q05	809	-20.83	0.43	0.46	698	-20.83	0.51	0.49	165	-2.26	0.16	0.22
q10	809	-88.64	0.62	0.69	698	-88.64	0.72	0.74	165	-3.07	0.26	0.26
q25	809	-1.54	1.10	1.35	698	-1.54	1.27	1.46	165	-1.20	0.45	0.51
q50	809	-0.16	1.86	2.18	698	-0.16	2.09	2.33	165	-0.16	0.80	0.85
(b) Income surveys												
IPL	1017	-7385	0.11	16.91	905	-7385	0.22	19.08	3	0.22	0.23	0.24
NPL	1017	-3248	0.50	6.11	905	-3248	0.51	6.34	3	-	-	-
SPL	1017	-3453	0.41	16.36	905	-3453	0.48	18.38	3	0.23	0.27	0.32
q05	1017	-5077	-0.06	1.86	905	-5077	-0.01	2.20	3	-0.13	-0.06	-0.05
q10	1017	-2777	-0.01	14.92	905	-2777	0.07	16.81	3	0.00	0.03	0.05
q25	1017	-1894	0.25	19.61	905	-1894	0.35	22.04	3	0.01	0.02	0.10
q50	1017	-1880	1.71	-0.25	905	-1880	1.98	-0.39	3	0.10	0.13	0.36

Note: This table reports the values of the consumption floor using the Beta distribution as outlined in equation 2. Floors are reported in daily per capita 2017 PPP\$. See also Table 4.

Acronyms: z - poverty line; N – number of surveys; Min – minimum; p50 – median; IPL – International poverty line; NPL – national poverty line; SPL – societal poverty line; q05 – threshold income of 5<sup>th</sup> percentile; q10 – threshold income of the 10<sup>th</sup> percentile; q25 – threshold income of the 25<sup>th</sup> percentile; q50 – threshold income of the 50<sup>th</sup> percentile.

Table Dxx: Share of population below \$0.25 and \$0.50 per day, 10 most affected surveys

Rank	Below \$0.25/day						Below \$0.50/day					
	Consumption			Income			Consumption			Income		
	Country	Year	Pop share, %	Country	Year	Pop share, %	Country	Year	Pop share, %	Country	Year	Pop share, %
1	Zambia	1991	15.93	Suriname	1999	11.04	Dem. Rep. Congo	2004	30.22	Haiti	2001	16.54
2	Lesotho	1994	7.14	Panama	1989	9.79	Zambia	1991	25.39	Haiti	2012	15.11
3	Lesotho	2002	6.14	Colombia	2000	7.99	Mali	1994	24.04	Micronesia	2000	13.05
4	Dem. Rep. Congo	2004	6.12	Haiti	2001	7.74	Mozambique	1996	22.33	Panama	1989	11.87
5	Mozambique	1996	4.76	Panama	1991	6.98	Lesotho	1994	16.59	Suriname	1999	11.81
6	Mali	1994	4.68	Haiti	2012	6.95	Niger	1994	16.42	Bolivia	2000	10.20
7	Uzbekistan	1998	4.29	Panama	1995	6.70	Lesotho	2002	16.27	Panama	1991	8.99
8	The Gambia	1998	3.57	Micronesia	2000	5.82	Burkina Faso	1994	12.92	Colombia	2000	8.53
9	Burundi	1998	3.42	Panama	1998	5.30	The Gambia	1998	12.53	Bolivia	1999	8.40
10	Benini	2015	3.00	Colombia	1999	5.20	Burundi	1998	12.41	Nicaragua	1993	8.38

Note: This table reports the top 10 income and consumption surveys and the share of population that are below \$0.25 per day and \$0.50 per day thresholds from the most affected (rank 1) to the least.