development
initiatives

December 2018

# filling the gaps in current global poverty data estimates

discussion paper

# Contents

# Summary

As the authoritative global repository of extreme poverty data, the World Bank's PovcalNet database currently provides contemporary and historical poverty estimates for 164 economies.[1] Despite this excellent coverage, we have identified 77 economies[2] (18 of which are ODA eligible) with no available data for the most recent reference year of 2015. This is a data gap for almost 200 million people worldwide, which may only represent 2.7% of global population but, as many of these people live in poor countries when measured by other standards (e.g. revenues per capita) and in hard-to-reach places, they may represent a much larger relative percentage of the global population living in extreme poverty. There is, therefore, significant interest in estimating the state of poverty in these places to ensure that no one is left behind.

In this paper, firstly we assess the possibility of filling these data gaps in a way that builds on other commonly used methods of estimation. Secondly, we produce poverty estimates for five missing data economies, covering almost 60 million people, using the methods proposed. The innovative approaches used here seek to improve analytical and decision-making value of poverty estimates by providing a new opportunity to evaluate the characteristics of extreme poverty in some of the world's uncounted poor.

The methods presented in this paper are not intended to create infallible estimates of poverty, and our results do not stand up to the same rigour as those in the PovcalNet databank; as such, we welcome suggestions of improvement to this methodology. Our results identify 25.2 million people in extreme poverty in the five chosen missing data economies, representing an increase of 11.6 million people over estimates based on their respective regional averages. Notably, our results are in line with expected poverty headcounts based on complementary analyses in these missing economies and go beyond single values to allow for further calculation of other useful statistics, including distributional information. We therefore believe that our estimates are a better alternative to other methods used to estimate poverty in uncounted populations.

# Current data gaps

In 2015, world leaders agreed a set of Sustainable Development Goals (SDGs), the first goal of which is a commitment to ending extreme poverty 'for all people everywhere' by 2030. The presence of gaps in poverty data is a stumbling block for analysts and policymakers meeting this goal.

As a leader in the fight against poverty, the World Bank stands also as the global authority for international poverty data. The World Bank's PovcalNet repository publishes biennial estimates of poverty and welfare distributions based on over 1,500 household surveys, spanning 1979 to 2017. PovcalNet is thus an integral tool for development actors, providing contemporary and historical estimates for extreme poverty data in 164 economies. However, despite constantly improving coverage, PovcalNet does not yet provide a truly comprehensive overview of global poverty: we have identified 77 economies or territories[3] with no available current extreme poverty data; this represents 197 million people, 117 million of whom are situated in official development assistance (ODA)-eligible recipient nations. This lack of complete coverage highlights a fundamental dilemma, namely that PovcalNet's rightly rigorous data requirements can prevent the publishing of poverty data where data is incomplete or of poor quality.

This gap, although small on a global scale (2.7% of world's population[4]), are particularly concentrated in regions of fragility or in countries at risk of being left behind: 85 million people not covered by existing poverty data live in fragile states, and 52 million people live in countries at risk of being left behind (CBLB).[5] Those people uncounted by current poverty estimates are therefore some of the most difficult-to-reach people and may represent a much larger relative percentage of the global population living in extreme poverty.

Methods for accommodating the lack of poverty data vary with no commonly accepted approach. The immediate solution used by PovcalNet is to apply respective regional averages to the missing population while redacting the missing data from tabular results. In other cases, populations with missing data may be excluded from analyses, but rendered as '*No data*' in tables or charts. Neither approach provides much meaningful information for policymakers, and at worst can convey incorrect messaging.

Alternatively, an attempt to estimate poverty through regression-based approaches may be made: for example, estimating the relationship between macro-level economic variables and extreme poverty. One of the most prominent methods in estimating missing poverty data is the approach used by the World Poverty Clock, which uses a logistic regression based on GDP per capita and a dummy variable for oil-exporting economies. However, as will be demonstrated in our report *Alternative methods for filling gaps in poverty data*, which will be published in early 2019, such a model is unlikely to consider all relevant variables and is in many cases found to be a poor predictor of extreme

poverty. Furthermore, any such regression-based method provides only a single resultant value for estimated poverty headcount and does not provide the ability to calculate other useful statistics such as Gini index, Foster–Greer–Thorbecke indices,[6] Sen index or other distributional information.

This paper summarises a method for filling poverty data gaps, seeking to improve on current methods of estimation by mirroring the process used by PovcalNet as described by Chen and Ravallion (2010),[7] among others. This is achieved using household-level information where available and proxy or secondary sources otherwise. The result is a robust and comprehensive poverty estimate with the further ability to calculate other useful statistics.

### Table 1: ODA-eligible economies with no PovcalNet poverty data

| Economy | Population | Is it fragile? | Is it being left behind? |
|---|---|---|---|
| Afghanistan | 33.7m | Yes | Yes |
| Democratic People's Republic of Korea | 25.2m | Yes | |
| Cambodia | 15.5m | | |
| Somalia | 13.9m | Yes | Yes |
| Cuba | 11.5m | | |
| Libya | 6.2m | Yes | |
| Eritrea | 4.8m | Yes | Yes |
| Argentina (rural areas) | 3.7m | | |
| Equatorial Guinea | 1.2m | Yes | |
| *9 others with populations under 1m\** | *~1.5m* | | |

Source: Author's calculations based on Development Initiatives, PovcalNet, World Bank World DataBank, Organisation for Economic Co-operation and Development (OECD) States of Fragility 2018 and United Nations Population Division.

Notes: Population data are for 2015; fragile indicator is based on OECD States of Fragility 2018; countries being left behind are based on Development Initiatives, 2018.[8] *Macao (special administrative region, China), Western Sahara, St Vincent and the Grenadines, Grenada, Antigua and Barbuda, Dominica, Marshall Islands, Palau, Nauru.

# Why gaps exist

To identify why data gaps occur, we need to understand the process through which poverty estimates are calculated by PovcalNet. Three primary data components are required to generate PovcalNet's results for each individual economy: (1) the population's mean consumption/income, (2) the population consumption/income distribution and (3) the private consumption purchasing power parity (PPP) conversion factor. PovcalNet derives each of these primary components directly from representative surveys, national accounts and the International Comparison Program (ICP).[9]

**Figure 1: Process for estimating poverty from primary data sources**
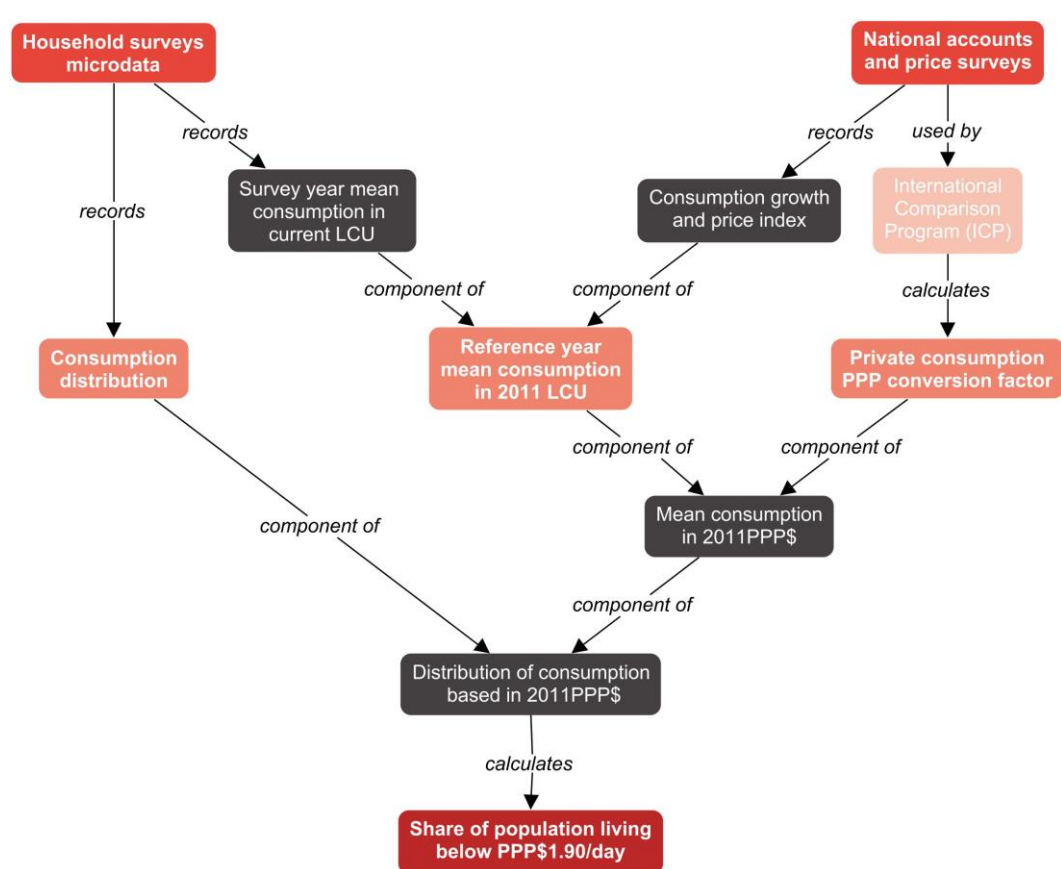


Notes: Process is simplified for illustrative purposes. 2011 LCU is local currency units in 2011 prices; 2011PPP$ is purchasing power parity dollars in 2011 prices.

In the event a survey has not been undertaken, is not representative of a population, or faces other reliability issues, then a poverty estimate using PovcalNet's process is not rendered. The various reasons for current gaps in poverty data therefore stem from two

core issues: data availability and data confidence. In the former, there is a substantial lack of applicable data from which poverty estimates can be derived. In the latter, there exists applicable data, but questions as to its quality or veracity prevent a confident poverty estimate from being calculated: in many cases, missing components are indicative of PovcalNet's stringent data requirements for the underlying data. It is therefore unsurprising that difficult-to-reach economies and territories are most likely to have data gaps.

A third issue that may hinder poverty estimates with PovcalNet data is the frequency of data release – PovcalNet updates its repository twice a year and will only present new data once every other year. In some cases, poverty data may be newly available but not yet incorporated into PovcalNet's database; any new data must be comprehensively assessed by World Bank economists before being used in a PovcalNet update.

Component gaps for ODA-eligible economies with populations over 1 million people missing current poverty data are identified in Table 2. Component availability was assessed based on their availability from the PovcalNet repository, World Bank Microdata Library and ICP repository.

**Table 2: Component availability of selected ODA-eligible economies missing poverty data**

| Economy | Mean | Distribution | PPP conversion |
|---|---|---|---|
| Afghanistan | Not available | Available | Not available** |
| Democratic People's Republic of Korea | Not available | Not available | Not available |
| Cambodia | Not available* | Available | Available |
| Somalia | Available | Available | Not available[†] |
| Cuba | Not available | Not available | Available |
| Libya | Not available | Not available | Not available** |
| Eritrea | Not available | Not available | Not available** |
| Argentina (rural areas) | Not available | Not available | Not available |
| Equatorial Guinea | Not available | Not available | Available |

Source: Author's calculations based on ICP, PovcalNet and World Bank Microdata Library.

Notes: *Cambodia's welfare aggregates are currently considered unreliable,[10] **PPP conversions are available as non-benchmark estimates; [†]Somalia's poverty-specific PPP conversion is indirectly available from Somali High Frequency Survey.[11] Mean and distribution components are considered available to PovcalNet if they are present in the PovcalNet repository (2005 or 2011 version) or World Bank Microdata Library; PPP conversion component is considered available if it is present in the ICP 2011 repository.

## Box 1: Primary components of poverty estimates

The three primary components used in an extreme poverty estimate are a population's consumption/income mean, distribution and private PPP conversion. These three components are each required to generate a representative distribution based in PPP$. Applying the extreme poverty threshold of $1.90 a day reveals an estimate for the proportion of population living below the international extreme poverty line.

PovcalNet's core analysis builds on representative household surveys, which record unit consumption/income data of representative populations. Mean consumption/incomes and prices may be aggregated at either a subnational level, allowing for within-economy spatial-price adjustments, or national level. From this information, the mean is derived as a real monetary value in local currency units.

In most cases, the distribution is also derived directly from the available household survey microdata responses, wherein a welfare aggregate will be generated and from which an accurate discrete distribution may be generated. In other cases where microdata is unavailable, distributions are derived from an aggregation of responses and presented as a set of quantile values. In both cases, these data are generally presented in the form *L(P)*, that is, the cumulative quantile of consumption/income *(L)* held by the bottom *(P)* cumulative quantile of population – the basis of a Lorenz curve. These quantile values allow a parametric estimation of the welfare distribution to be made as per Datt in 1998.[12]

The private consumption PPP conversion factor is calculated in a separate process by the ICP, which uses reported local prices, national accounts and other sources to derive an equivalent PPP conversion for an economy. In some cases, PovcalNet adjusts the resulting applied PPP conversion factor based on subnational differences as per Ferreira et al in 2016.[13] This conversion factor is necessary to evaluate local currency at an equal purchasing power to the US dollar within the US.

# Filling the gaps

Here, we seek to parallel PovcalNet's process of estimating poverty headcounts as closely as possible. We have further chosen to 'line up' our estimates to PovcalNet's most recent refence year, which is 2015. The underlying methodology used is covered extensively in Chen and Ravallion (2010),[14] Datt (1998),[15] Ferreira, et al. (2016)[16] and World Bank (2018),[17] which, while not without its caveats or limitations, represents the best approach in estimating global poverty.

To demonstrate the methodology, we have selected five ODA-eligible economies missing poverty data. Those selected are: Afghanistan, Somalia, Libya, Eritrea and Equatorial Guinea. Together, these economies have 59.9 million people – representing over half the number of people not currently included in poverty estimates in ODA-eligible economies. Having identified the missing primary components for our example economies, it is necessary to find appropriate alternatives. Secondary sources, such as unofficial data, external estimates or proxies are potential solutions, but they must be individually evaluated to ensure their suitability for the application. We selected suitable secondary sources for the identified missing components as follows:

**Mean:** A population's mean consumption/income, where not immediately available for the 2015 reference year, was projected from either a non-2015 survey value or the 2011 ICP estimate based on the methodology described by World Bank (2015).[18] Where necessary, the mean was deflated to 2011 prices using the official consumer price index (CPI) and converted to PPP$ using the appropriate conversion factor (see below).

**Distribution:** Where the distribution of consumption/income was not available from PovcalNet's repository, other official sources were considered, such as socio-economic or living conditions surveys; where used, data was drawn from national statistics offices. Where no accessible consumption/income distribution was found, a proxy dataset was used – the primary choice for this was the relative wealth distribution as recorded by Demographic and Health (DHS) surveys and Multiple Indicator Cluster surveys (MICS). Despite not measuring consumption/income directly, wealth distributions provide a good approximation for relative quantile inequalities. Our forthcoming report *Alternative methods for filling gaps in poverty data*, to be published in early 2019, will outline the complete methodology and comparisons between poverty estimates made using consumption/income and wealth distributions for selected economies.

**PPP conversion:** The primary source for official PPP conversion data was the International Comparison Program (ICP); additionally, in limited cases, conversions were drawn from national sources. If official conversions were unavailable, where possible a regression-based estimate was used (this is described in our forthcoming

report *Alternative methods for filling gaps in poverty data*). Where neither official estimates nor regression-based estimate were attainable, ICP non-benchmark estimates were used.

**Table 3: Component sources of selected example economies**

| Economy | Mean | Distribution | PPP conversion |
|---|---|---|---|
| Afghanistan | Afghanistan Living Conditions Survey 2016–2017 || Development Initiatives estimate |
| Somalia | Somali High Frequency Survey 2016–2017 ||| 
| Libya | Libya National Socio Economic Survey 2008 || ICP 2011 (non-benchmark estimate) |
| Eritrea | ICP 2011 (non-benchmark estimate) | Eritrea DHS 2002 | Development Initiatives estimate |
| Equatorial Guinea | Equatorial Guinea Poverty Profile 2006 | Equatorial Guinea DHS 2011 | ICP 2011 |

Source: See *Data Sources.*

## Box 2: Parametric Lorenz curves in poverty estimation

PovcalNet's poverty estimation methodology relies on either household (unit)-level or grouped data for its distribution component. In the former case, poverty estimates are directly rendered by the proportion of observed households below the poverty line; in the latter case, it is necessary to model the distribution with a parametric Lorenz curve.

Two models of parametric Lorenz curve are used by PovcalNet – the Beta Lorenz curve and the general quadratic Lorenz curve. Both models are demonstrated to perform well at fitting grouped distribution data.[19] The choice of specification is determined by model validity and goodness of fit.[20,21] From a fitted parametric Lorenz curve, a poverty estimate for any given reference year is retrieved by applying the appropriate mean and solving the specification at the given poverty line. This approach is used in all cases where non-unit-level data is available.

# Results and discussion

The chosen example countries have 59.9 million people that are currently uncounted – our results suggest 25.2 million of this group live below the international extreme poverty line of $1.90 per day. The highest estimated poverty headcount ratios occur in Somalia (71.7%) and Eritrea (50.9%), both above their respective regional average of 41.1%. Our results also return higher than regional averages for Afghanistan and Libya. Only Equatorial Guinea has an estimated poverty headcount ratio lower than its respective regional average. Based on these results, the global total of people living in extreme poverty would be revised upwards by 11.6 million[22] people.

**Table 4: Estimated 2015 extreme poverty headcount % (and number of people living in poverty)**

| Economy | Regional average | World Poverty Clock | Development Initiatives estimate |
|---|---|---|---|
| Afghanistan | 15.1% (5.1m) | 39.0% (13.1m) | **35.2% (11.9m)** |
| Somalia | 41.1% (5.7m) | 52.6% (7.3m) | **71.7% (10.0m)** |
| Libya | 5.0% (0.3m) | <3% (<0.1m) | **8.1% (0.5m)** |
| Eritrea | 41.1% (2.0m) | 43.3% (2.1m) | **50.9% (2.5m)** |
| Equatorial Guinea | 41.1% (0.5m) | <3% (<0.1m) | **29.3% (0.3m)** |

Source: Author's calculations based on the Central Intelligence Agency World Factbook, PovcalNet, World Bank DataBank and World Poverty Clock.

Notes: World Poverty Clock data is not available for 2015; values are therefore based on the methodology and data parameters described by Cuaresma et al (2018).[23]

As with all poverty estimates, our results rely on the integrity of the chosen primary components. For example, in cases where regression-based PPP conversions have been used, such estimates may not be as accurate as official figures. This is further

compounded by the expected significant spatial price differences within economies experiencing conflict. Despite this limitation, we ultimately feel that the methodology derived in our forthcoming report *Alternative methods for filling gaps in poverty data* currently presents the best alternative to official ICP benchmark estimates.

A second limitation is in estimates that have used a relative wealth distribution. Wealth and consumption/income are not the same concept, as such there is no guarantee that the relative wealth index always provides a suitable proxy for the economy consumption/income distribution. Despite this, our analysis demonstrates that wealth distributions are often a close analogue for consumption distributions; closer even, it seems, than income is an analogue to consumption.[24] Therefore, and with note of PovcalNet's conflation of consumption and income, we feel that the use of relative wealth distributions is not unreasonable and provides the next best alternative where consumption/income distributions are not available.

Finally, we acknowledge that the use of parametric Lorenz curves from wide-grouped distributional data is substandard to unit-level analysis, particularly in cases where the poverty line falls near the tails of the distribution. This issue is alieved by higher resolution grouping or, ultimately, the availability of unit-level data. The absence of such data makes our estimates, to a degree, reliant on the parametric specification's characteristics. However, this consideration is not unique to our methodology, and is itself present in PovcalNet's estimates where they are based on grouped data. We therefore feel comfortable with the approach used.

Beyond methodological concerns, we have examined our individual results for parity with other measures: for example, Afghanistan's result of 35.2% may superficially appear low, particularly for a low-income economy emerging from decades of disruptive conflict. However, a complementary analysis of the Multidimensional Poverty Index demonstrates that our estimate is in line with non-monetary indicators of deprivation. Furthermore, a comparison with Afghanistan's nationally defined poverty thresholds places our estimate of extreme poverty reasonably above the food poverty threshold, and below the national poverty threshold. For further robustness analysis, including all estimated economies, see our forthcoming report, *Alternative methods for filling gaps in poverty data*, to be published in early 2019.

Despite limitations, the results presented here demonstrate the usefulness of available data on uncounted people living in poverty in making extreme poverty estimates. Until official estimates are otherwise available, innovative and unconventional approaches (such as those presented here) represent a potentially valuable resource for development actors in ensuring that no one is left behind in the fight against poverty.

# Annex

**Table A1: Economies with no PovcalNet poverty data, grouped by World Bank region (economies with population over 1 million are shown in bold)**

| Region | Economies |
|---|---|
| East Asia and Pacific | American Samoa, Brunei Darussalam, **Cambodia**, Christmas Island, Cocos (Keeling) Islands, Cook Islands, French Polynesia, Guam, **Hong Kong (SAR, China)**, **Democratic People's Republic of Korea**, Macao (SAR, China), Marshall Islands, Pitcairn Islands, Nauru, New Caledonia, **New Zealand**, Niue, Norfolk Island, Northern Mariana Islands, Palau, **Singapore**, Tokelau |
| Middle East and North Africa | **Bahrain**, **Kuwait**, **Libya**, **Oman**, **Qatar**, **Saudi Arabia**, **United Arab Emirates**, Western Sahara |
| South Asia | **Afghanistan** |
| Latin America and the Caribbean | Anguilla, Antigua and Barbuda, **Argentina (rural areas)**, Aruba, the Bahamas, Barbados, Bermuda, British Virgin Islands, Cayman Islands, **Cuba**, Curacao, Dominica, Falkland Islands, French Guiana, Grenada, Guadeloupe, Martinique, Montserrat, **Puerto Rico**, Sint Maarten (Dutch part), Saint Barthélemy, Saint Pierre and Miquelon, Saint Kitts and Nevis, Saint Martin (French part), Saint Vincent and the Grenadines, Turks and Caicos Islands, Virgin Islands (US), Wallis and Futuna |
| Sub-Saharan Africa | **Equatorial Guinea**, **Eritrea**, Reúnion, **Somalia**, Saint Helena, Ascension and Tristan de Cunha |
| Europe and Central Asia | Åland Islands, Akrotiri and Dhekelia, Andorra, Channel Islands, Faroe Islands, Gibraltar, Greenland, Isle of Man, Liechtenstein, Malta, Monaco, San Marino, Vatican City |

Notes: The designations used do not imply the expression of any opinion on the part of Development Initiatives concerning the legal status of a territory or of its authorities. SAR, special administrative region.

# Data sources

Afghanistan Living Conditions Survey 2016–2017, Afghanistan Central Statistics Organisation. Available from: http://cso.gov.af/en/page/1500/1494/nrav-report (accessed October 2018)

Equatorial Guinea Poverty Profile 2006, AFRISTAT/DGSCN. Available from: http://catalog.ihsn.org/index.php/catalog/6815/ (accessed October 2018).

Eritrea DHS 2002 Wealth Index, The DHS Program. Available from: https://www.dhsprogram.com/topics/wealth-index/Wealth-Index-Construction.cfm (accessed October 2018).

Libya National Socio Economic Survey 2008, Libyan Bureau of Statistics and Census. Available from: http://bsc.ly/ (accessed October 2018).

International Comparison Program (ICP), World Bank. Available from: http://www.worldbank.org/en/programs/icp#5 (accessed October 2018).

Somali High Frequency Survey Wave 2, Somali Directorate of National Statistics. Available from: http://www.dns.org.so/microdata/index.php/catalog/13 (accessed October 2018).

PovcalNet, World Bank. Available from: http://iresearch.worldbank.org/PovcalNet/povDuplicateWB.aspx (accessed October 2018).

World Bank DataBank, World Bank. Available from: https://data.worldbank.org/ (accessed October 2018).

The World Factbook, Central Intelligence Agency. Available from: https://www.cia.gov/library/publications/the-world-factbook/ (accessed October 2018).

# Notes

[1] 163 World Bank economies, plus Argentina (urban areas).

[2] 53 World Bank economies, plus Argentina (rural areas), Western Sahara and 20 non-World Bank territories.

[3] Beyond sovereign states, this list also includes non-sovereign, disputed, dependent and non-self-governing territories that are not represented by current poverty figures. For a full list of territories covered, see Annex.

[4] This coverage gap is based on economies where no data is presented by PovcalNet. Other data gaps may occur where surveys are not representative of an entire economy population. PovcalNet records a figure of 'regional survey coverage', which is the total share of population of a region that is covered by a representative survey, and therefore excludes both types of data gap.

[5] For more information on the 'countries being left behind' concept, see Countries being left behind, available at: http://devinit.org/post/countries-left-behind/

[6] The Foster–Greer–Thorbecke indices are commonly presented as: FGT(0) – headcount ratio, FGT(1) – poverty gap index, and FGT(2) – squared poverty gap index.

[7] Chen S. and Ravallion M. The developing world is poorer than we thought, but no less successful in the fight against poverty. The Quarterly Journal of Economics, 125:4, 2010, p. 1577–1625.

[8] Development Initiatives, 2018. Countries being left behind, available at: http://devinit.org/post/countries-left-behind/

[9] World Bank, 2017. Monitoring Global Poverty: Report of the Commision on Global Poverty. Washington, DC: World Bank.

[10] World Bank, 2017. What is new 2017/10/10. Available at: http://iresearch.worldbank.org/PovcalNet/WhatIsNew.aspx

[11] World Bank, 2017. Somali poverty profile: findings from wave 1 of the Somali high frequency survey (English). Available at: http://documents.worldbank.org/curated/en/325991506114032755/pdf/AUS19442-REVISED-PUBLIC-Somali-Poverty-Profile-Update-4-27-November-2017.pdf (accessed October 2018).

[12] Datt G. Computational Tools for Poverty Measurement and Analysis. Food Consumption and Nutrition Division Discussion Paper No. 50, 1998, International Food Policy Research Institute.

[13] Ferreira, F.H., et al. A global count of the extreme poor in 2012: data issues, methodology and initial results. The Journal of Economic Inequality, 14:2, 2016, p. 141–172.

[14] Chen S. and Ravallion M. The developing world is poorer than we thought, but no less successful in the fight against poverty. The Quarterly Journal of Economics, 125:4, 2010, p. 1577–1625.

[15] Datt G. Computational Tools for Poverty Measurement and Analysis. Food Consumption and Nutrition Division Discussion Paper No. 50, 1998, International Food Policy Research Institute.

[16] Ferreira, F.H., et al. A global count of the extreme poor in 2012: data issues, methodology and initial results. The Journal of Economic Inequality, 14:2, 2016, p. 141–172.

[17] World Bank, 2018. PovcalNet Methodology. Available at: http://iresearch.worldbank.org/PovcalNet/methodology.aspx

[18] World Bank, 2015. A Measured Approach to Ending Poverty and Boosting Shared Prosperity: Concepts, Data, and the Twin Goals. Washington, DC: World Bank.

[19] Chen S., et al. Is poverty increasing in the developing world? Review of Income and Wealth, 40:4, 1994.

[20] Kakwani N. and Podder N. Efficient Estimation of the Lorenz Curve and Associated Inequality Measures from Grouped Observations. Econometrica, 44:1, 1976, p. 1–22.

[21] Villaseñor J. and Arnold B.C. Elliptical Lorenz curves. Journal of Econometrics, 40:2, 1989, p. 327–338.

[22] PovcalNet indirectly renders 13.6 million as the total people in extreme poverty in these missing economies through regional averages.

[23] Cuaresma J.C., et al. Will the Sustainable Development Goals be fulfilled? Assessing present and future global poverty. Palgrave Communications, 2018, Article number: 29.

[24] Development Initiatives (forthcoming). *Alternative methods for filling gaps in poverty data.*