Is Inequality Systematically Underestimated in Sub-Saharan Africa?

A Proposal to Overcome the Problem

Fabio Clementi
Michele Fabiani
Vasco Molini
Francesco Schettino

Abstract

In Africa, evidence on the interactions among poverty, growth, and income distribution presents a puzzle: While growth has been robust in recent decades, the growth elasticity of poverty has remained low. This suggests that inequality has dampened the pro-poor effects of growth. However, when using standard inequality measures, there is only scattered evidence of high and growing inequality in Africa outside the extremely unequal southern cone. This paper argues that inequality mismeasurement could be the main culprit responsible for this paradox: consumption-based measures miss important information at the top end of the consumption distribution, leading to underestimation of inequality. This paper proposes distinct solutions, arguing that by reevaluating the importance of distributional issues in Africa, the need becomes apparent for refreshing the research agenda on African development in such a way that the interaction between poverty and inequality becomes a core concern.

Is Inequality Systematically Underestimated in Sub-Saharan Africa? A Proposal to Overcome the Problem

Fabio Clementi

University of Macerata, Macerata, Italy

Michele Fabiani

University of Campania "Luigi Vanvitelli", Caserta, Italy, and University of Macerata, Macerata, Italy

Vasco Molini[1]

The World Bank, Rome, Morocco

Francesco Schettino

University of Campania "Luigi Vanvitelli", Caserta, Italy

---

[1] *Corresponding author*. E-mail: vmolini@worldbank.org

1   Introduction

Over the past twenty years, Sub-Saharan Africa (SSA) has experienced an unprecedented resurgence of economic growth. While this growth is encouraging for the prospects of SSA's economic development, there is an ongoing debate regarding its nature and outcomes (*inter alia* Fosu 2009, 2017a, 2017b, 2018; Christiansen et al. 2013; Harttgen et al. 2013; Cornia 2017; Odusola et al. 2017). While most countries in SSA have experienced reductions in poverty, the progress has been relatively slow compared to non-African developing countries experiencing similar growth rates (Thorbecke and Ouyang 2018). An intuitive explanation for this suboptimal performance is given by the abundant literature on growth non-inclusiveness in SSA (*inter alia* Christiansen et al. 2013; Cornia 2017; Odusola et al. 2017). Several economists working on postcolonial SSA have argued that the growth process was described as a rent-seeking one, driven by rising rents from resource extraction, as the form and function of extractive institutions from a colonial past have been maintained (Devarajan and Giugal 2013; Atkinson 2014; Knight Frank Research 2015; Adhvaryu et al. 2021). In a nutshell, economic growth, when driven by a resource boom and presided over by extractive institutions, disproportionately benefits a country's ruling elite rather than the poor (Spinesi 2009; Acemoglu and Robinson 2012; Sala-i-Martin and Subramanian 2003; Robinson et al. 2006; Devarajan and Giugal 2013). Consequently, many have observed that the growth pattern in SSA has remained unevenly distributed and has largely failed to "trickle down" to the poor.

We would therefore expect that high and increasing inequality is the principal culprit for SSA's comparatively poor record in translating growth into poverty reduction over the past twenty years. The fact that the gross domestic product (GPD) growth experienced in past decades did not accompany a commensurate reduction in poverty would suggest a generalized increase in inequality in these countries as postulated by the so-called poverty-inequality-growth triangle (Bourguignon 2004). However, the evidence for this is rather scattered and ambiguous. SSA is frequently said to rival Latin America as the most unequal region in the world, but this aspect seems to be mainly driven by the few exceptionally unequal countries in Africa's Southern Cone (Odusola et al. 2017, Clementi et al., 2020). Excluding the Southern Cone, inequality in SSA is not high by developing-country standards. Evidence of recent trends in inequality during the SSA's growth "miracle" is also mixed—no clear pattern emerges that could hold generally across the continent (Pinkovskiy and Sala-i-Martin 2014; Beegle et al. 2016; Odusola et al. 2017).

The inequality performance in SSA also raises questions regarding development theory, the application of which would predict a systematic increase in inequality in SSA over the past two decades (Spilimbergo et al. 1999; Beegle et al. 2016). Literature on the topic (*inter alia* Lewis 1955; Gollin et al. 2014) has noted that, even during the recent economic take-off, large productivity gaps between agriculture and nonagricultural sectors have persisted (MacMillan et al. 2014). According to the Kuznets curve theory (Kuznets 1955; Kanbur 2017), inequality first increases and then declines as a country develops. In SSA, the low initial level of development as well as persistent sectoral and spatial productivity gaps would predict an increase in income differentials during the early growth spurt. However, as mentioned, inequality figures from SSA fail to provide clear evidence for a Kuznets-type trajectory (Ravallion 2005). Also, the existing significant positive relationship between the level of GDP and inequality (Bhorat et al. 2016; Beegle et al. 2016) is in fact driven almost entirely by the Southern Cone.

This puzzle induced scholars like Jerven (2015), to question the veracity of numbers and accuracy of measures used. On one side, many have questioned the extent to which the GDP growth has been driven by structural transformation of African economies. They maintain that much of the recent performance seems to be due to temporary boosts: advantageous external context and making up of lost ground after a long period of economic decline. In other words, describing a growth model mainly due to fortuitous and potentially temporary favorable external conditions (Rodrik 2016a; Bhorat et al. 2017; Diao et al. 2018), they underline that the traditional engines behind rapid GDP growth, structural change and industrialization operated at less than full power, implicitly resizing the extent of the African miracle, especially from a qualitative viewpoint.

All of these explanations have several shortcomings. First, they do not take into adequate account the fact that inequality as it is currently measured in SSA could be simply underestimated. The limitations of using consumption as the widely preferred welfare measure to investigate inequality in SSA, we argue, could be partly able to explain this conundrum. Our belief is that consumption as a proxy for well-being is well-suited to measuring poverty or in general the well-being of the bottom deciles (Meyer and Sullivan 2011), but its appropriateness as a proxy for well-being fades as one moves up the consumption distribution (Chancel et al. 2019). This is because the basket of goods and services commanded by top deciles is not well represented in standard consumption surveys, which, after all, are designed primarily for poverty measurement and hence focus on a fairly basic basket of goods.

Failing to fully capture the consumption of the middle and upper classes—in SSA they occupy the top two deciles (African Development Bank 2011; Shimeles and Ncube 2015; Corral Rodas et al. 2019)—may lead to an underestimation of the welfare of this group. While this does not pose a problem for poverty measurement, it does mean that consumption is less well suited to measuring inequality. Furthermore, this problem is exacerbated when attempts are made to measure inequality changes over time, since the creation or consolidation of national middle classes in SSA pushes a greater proportion of households into that part of the distribution that is poorly captured by consumption measures (Kharas 2010; Ravallion 2010; Ncube and Lufumpa 2014; Schotte et al. 2018)—potentially leading to an underestimation of the rise of inequality. This, we argue, calls for adjusting the way we measure welfare in the region.

Our proposal to overcome this problem is to adjust the top of the consumption distribution by using in-sample information from a corresponding income distribution (Clementi et al. 2020). In a nutshell, our approach is to recalibrate the consumption figures for the middle-class segments by imputing information from the shape of the income distribution of the same households. This recalibration of the consumption distribution for a sample of selected African countries is made possible by constructing an ad hoc database that combines information on consumption with information obtained from household-based surveys maintained by the Food and Agriculture Organization (FAO) of the United Nations, which provides several important indicators on rural livelihoods but also constructs income estimates at the household level.

The paper is organized as follows. Section 2 details the data and the methodology used to produce our results. Section 3 presents unequivocal evidence of inequality underestimation by means of multivariate statistics and econometric analysis. Section 4 suggests original solutions to overcome the problem, also providing the empirical results. Section 5 concludes.

2   Data

Various data sources are used for this paper. The general overview on consumption versus income (section 3) is conducted using grouped data from PovcalNet, the global database of budget surveys that are conducted by national statistical offices under the supervision of government or international agencies

and collated by the World Bank.[2] These data include various measures of inequality and poverty, as well as a number of other useful distributional indicators, such as decile or (in some cases) percentile shares of the income or consumption distribution. We restrict our analysis to observations after 1980 and for which GDP and other basic socioeconomic variables are available. This reduces the sample to about 165 countries between 1980 and 2018 and includes all economic ranges, from low to high income.[3] The information from PovcalNet is combined with the World Development Indicators (WDI). This is the primary World Bank collection of development indicators compiled from officially recognized international sources.

There are more than 1,900 Gini and other inequality measures observations in our data set, most of which are calculated from direct access to household surveys. In the analysis, the Gini index—an indicator typically not sensitive to the tails of the income distribution—is accompanied by a measure that is more tail-sensitive, the D9/D5 ratio. A bigger gap between income and consumption inequality detected by this ratio is interpreted as a validation of our hypothesis that consumption-based measures tend to capture less accurately the information from the upper tails of the distributions.

In section 4, household-level data on income and consumption are taken from several sources. For a first group of SSA countries—Ghana, Kenya, Nigeria, and Uganda—we use household income data obtained from the Rural Income Generating Activities (RIGA) project, a collaborative effort of the FAO, the World Bank, and American University.[4] The RIGA database is composed of a series of constructed variables about rural and urban income-generating activities created from the original consumption data sources. In particular, we focus on the household-level income aggregate dataset (RIGA-H), which provides data on different income sources, such as crop and livestock production, household enterprises, wage employment, transfers, and nonlabor earnings.

For the countries just noted, data on households' consumption expenditure come instead from the original budget surveys compiled by national statistical bureaus and the World Bank, which can be easily linked

---

[2] World Bank, "PovcalNet: An Online Analysis Tool for Global Poverty Monitoring," database. http://iresearch.worldbank.org/PovcalNet.
[3] We follow the most recent (2020) World Bank classification that uses the Atlas method to estimate the size of economies in terms of gross national income (GNI) per capita: low income = below US$1,036; lower-middle income = between US$1,036 and US$4,045; upper-middle income = between US$4,046 and US$12,535; and high income = above US$12,535.
[4] The microdata are available to users upon request. For more see FAO, "The RIGA Database." http://www.fao.org/economic/riga/riga-database/en.

to each country's data set in the RIGA database. Specifically, as in Clementi et al. (2020), we consider here the following household budget surveys: the Ghana Living Standards Survey of 2005; the Kenya Integrated Household Budget Survey of 2005; the Nigeria Living Standards Survey of 2004; and the Uganda National Household Survey of 2005.

For a second group of SSA countries—Burkina Faso, Ethiopia, Malawi, Niger, Rwanda, and Tanzania—we rely on data from the Rural Livelihoods Information System (RuLIS), a major database to access cross-country comparable data and information on household income and expenditures at the microlevel.[5] Started as a joint research project among FAO, the World Bank, and the International Fund for Agricultural Development, the RuLIS database builds on the methodology developed and adopted by the RIGA project, whose procedures were deeply revised, integrated, and extended to RuLIS. In general, while the RIGA project aimed at constructing comparable income measures from household surveys in order to provide annualized benchmark aggregates, which—despite differences in the quality of information in each survey—would be suitable for cross-country analyses (Carletto et al. 2007), RuLIS includes a more standardized database that is meant to be a tool for analysis and wider dissemination. The RuLIS surveys[6] considered in this study cover a diverse set of SSA countries,[7] some of which have more waves of data and give us the opportunity to gauge the trends.

The main consumption variable that is used in the paper is the household total annual expenditure on food and nonfood items. As for income, the household income aggregates and their components included in the RIGA/RuLIS database closely follow the definition given by the International Labour Organization (ILO 2003), which considers as income receipts those that (1) recur regularly, (2) contribute to current economic well-being, and (3) do not arise from a reduction in net worth (Carletto et al. 2007). These three criteria are embodied in each of the components of income; irregular payments such as lottery earnings or inheritances, investments, savings, and the value of durables are not included in the RIGA/RuLIS definition and measure of income. Furthermore, costs are also considered to ensure that

---

[5] See FAO, "RuLIS—Rural Livelihoods Information System," database. http://www.fao.org/in-action/rural-livelihoods-dataset-rulis/en/.

[6] Burkina Faso, Enquête Multisectorille Continue (2014/15); Ethiopia, Ethiopia Socioeconomic Survey (2013/14, 2015/16); Malawi, Integrated Household Survey (2013, 2017); Niger, National Survey on Household Living Conditions and Agriculture (2011, 2014); Rwanda, Integrated Household Living Conditions Survey (2013/14); and United Republic of Tanzania, National Panel Survey (2009, 2012/13).

[7] Notice that both the RIGA and the RuLIS projects cover more SSA countries and provide datasets that are sometimes more recent than those used in the present analysis. However, limited coverage of the population and issues of accuracy caused us to focus only on the countries (and years) mentioned in the text.

the final income aggregate is net of costs, as opposed to gross (which could overestimate the income a household has at its disposal). Taxes are the only cost that has been subtracted from gross income earned to create net income earned (Quiñones et al. 2009).

Before undertaking the empirical analysis, both the income and consumption variables of each of the countries included in our sample were converted into constant 2011 international dollars—in purchasing power parity (PPP)—and expressed per capita per day. Furthermore, observations with negative and zero incomes were excluded from the analysis, because some indices of inequality are defined only for positive values. Accordingly, the sampling weights of households—used in all calculations—have been recalibrated in such a way that estimates from the samples after deletion of nonpositive records are forced to fit the initial population-level information on the households' geographical location and area of residence (rural versus urban).[8]

Table 1 presents distributional summary statistics for the survey-based consumption and income variables used in this study. Compared to income, consumption expenditure typically produces lower estimates of inequality, independently of the measure that one considers—the Gini coefficient, the mean log deviation (MLD), or the Theil index. As mentioned in the previous section, this is to be expected and can be explained by a declining marginal propensity to consume and by the fact that consumption surveys tend to understate the spending at the top. Instead, an argument for using consumption rather than income is that data on the former are often of a higher quality in developing and emerging economies and are less vulnerable to idiosyncratic shocks, as households tend to smooth their consumption over time. Because estimates of inequality will be biased if computed using any single one of these variables, what is needed to obtain consistent estimates of inequality, we argue, is a combination of the information coming from the consumption and income data. Section 4 appeals to multiple-imputation methods in order to achieve this.

---

[8] Calibration estimation, where the sampling weights are adjusted to make certain estimators match known population totals, is commonly used in survey sampling. Details on the subject can be found in, for example, Deville, Särndal, and Sautory (1993).

3 Income or Consumption?

3.1 Introduction

The surge in income inequality observed in the United States and other advanced economies in the past several decades (Jenkins et al. 2011; Piketty 2014, 2020; Schettino and Khan 2020) is stimulating an intense debate over whether other welfare measures confirmed or refuted the trend (Zaidi, and de Vos, 2001; Krueger and Perri 2006; Brzozowski et al. 2010; Jappelli and Pistaferri 2010; Fisher et al. 2015, 2016, 2021). From a theoretical point, consumption inequality received substantial attention since consumption is the closest proxy of utility among different welfare measures and because both the life-cycle hypothesis of Modigliani and Brumberg (1954, 1980) and the permanent income hypothesis of Friedman (1957) suggest that risk-averse households prefer a smooth to a variable consumption flow.

Consumption inequality can significantly differ from income inequality. Consumption may exceed income because a consumer is borrowing, but, at the same time, it may be below income because of the consumer's savings. Consumption can fall below income because of taxes paid, and, on the contrary, it can be higher, because of government transfers or home production, especially for households at the bottom of the distribution (Frazis and Stewart, 2011). Also, significant wealth effects can modify consumption behavior independently of income trend (Attanasio and Pistaferri 2016).

Specific considerations have to be added when we take into account developing country cases. Formal household monetary incomes are mostly constituted by wages and nonlabor monetary incomes (such as profits and rents). Most households in developing countries (for example, those in SSA), however, earn other forms of monetary incomes, such as those coming from agricultural production (both for selling and for autoconsumption) and from informal activities. In general, consumption is regarded as easier to measure than income in low-income economies, with its smoother behavior adding to the ease (Friedman 1957; Deaton and Zaidi 2002). Consumption can more adequately proxy permanent welfare, reducing the income short-run fluctuations; incomes from agriculture and informal activities routinely exhibit great seasonal variability (Tarozzi 2007).

When looking at inequality measures in developing countries, there are at least four reasons why those based on consumption tend to underestimate inequality compared to those based on income. First, consumption is more informative than income for the bottom of the distribution, since it reflects welfare, interpersonal transfers, and informal incomes (Meyer and Sullivan 2004), but it could underestimate

welfare of the top deciles of the distribution. This is principally due to the marginal propensity to consume that declines as household welfare increases (McCarthy 1995; Dynan et al. 2004; Jappelli and Pistaferri 2014; Gandelman 2017). Second, inequality calculated on consumption is likely to be biased downward if the set of goods in the consumption measure does not include items consumed by the rich (Beegle et al. 2016). Third, there may be households with zero annual income who, for example, finance their current spending out of previously accumulated savings, whereas people with zero annual consumption cannot exist. Finally, there are many income-rich people who save a part of their income; thus, their income can be greater than their consumption and, consequently, the high end of the distribution would be more elongated in the case of income (Milanovic 2010). This makes the distribution according to income more "elongated" both around the bottom and the top and thus more unequal: the consumption distribution will be "truncated" at some minimum amount necessary to survive and at the top since not all what households earn will be immediately spent.

Our empirical strategy, first of all, is to present the extent of the problem by conducting a multivariate analysis on a data set of developing and developed countries; this can show, all other things being equal, how much inequality is underestimated in SSA by means of using consumption. In the second part we propose our way to attain a better estimation of inequality in SSA. To our knowledge, this is the most reasonable way to correct the bias, since just substituting consumption with income would require a very expensive upgrade in the capacity of many national statistical offices in SSA and, second, because in many SSA countries' consumption is still the best indicator for measuring the welfare of many households whose revenues are from the informal sector and very seasonal (Clementi et al. 2020).

3.2   Empirical Estimations

Figure 1 illustrates this latter point using data from Povcalnet. High-income countries, although they still collect consumption data, tend to measure their official inequality figures with income; official inequality measured on consumption are practically nonexistent. Official consumption-based measures start to appear among the upper-middle income countries, although less numerous than those income-based, and eventually become the majority within the low-middle income group (figure 1, panel b). Among low-income countries, inequality is predominantly measured with consumption. Finally, looking at the different regions of the developing world, only in Latin America and the Caribbean do we have a predominance of inequality measured with income, while in other regions only the most advanced

countries use income—for example, in SSA only those upper-middle-income countries located in the Southern Cone use income data.

When in high-income countries we compare consumption to income inequality, this latter tends to be higher (Heathcote et al. 2010; Aguiar and Bils 2015); in other countries, due to data limitations, this comparison is not always feasible (Devarajan 2013; De Magalhães and Santaeulàlia-Llopis 2018). Nonetheless, a quick overview of Povcalnet enables us to identify 71 cases where inequality measures are available for the same country and same year both for consumption and income.[9] Figure 2 plots the differences in the Gini index (panel a) and the 90th percentile over the 50th percentile ratio (P90/P50-panel b) for these cases. In about 20 percent of the cases, the Gini based on income is slightly lower than that measured on consumption, but in the remaining 80 percent of the cases is higher. On average, income inequality is higher than consumption inequality by 10 percentage points, and in 40 percent of cases Gini measured on income exceeds that based on consumption by more than 10 percentage points. Interestingly, the gap between income and consumption slightly increases when looking at the P90/P50 ratio (panel b). In this latter case, the average gap between the two measures is 11 percentage points and about 30 percent of the observations have a gap higher than 20 percentage points. Indeed, the gap between the two measures tends to be more accentuated on the upper tail of the distribution rather than around the means.

In order to deeply inquire into this aspect, we estimate a model following the general formulation taken from the literature (Förster and Tóth 2015) that has enough explanatory variables but, at the same time, does not cover all the inequality determinants, because these are not easily available for many developing countries and for all the years we analyze.[10] Because the scope of this regression shows that—controlling for a set of socioeconomic variables and using different models—the use of consumption tends to underestimate inequality, our focus is mostly on the sign and significance of the consumption or income categorical variable.

---

[9] The countries are Bulgaria, Croatia, Estonia, Haiti, Hungary, Latvia, Lithuania, Mexico, Montenegro, Nicaragua, Philippines, Poland, Romania, Serbia, and Slovak Republic in different years.

[10] The data set used is the same constructed for plotting the figures but after eliminating the duplicates. Besides the mentioned cases where both income and consumption inequality measures are available—and we chose the one more used in the country in official publications—in some countries PovcalNet presents the inequality measures not only at the national level but also at the urban and rural levels and by subregions. In this case, we chose only the national measure. The final data set is an unbalanced panel covering 163 countries over almost four decades.

The generalized regression equation reads:

$$INEQ_{it} = \alpha + \delta\Gamma_{it} \times Inc\_Lev_{it} + \beta X_{it} + \gamma Y_{it} + REG_i + \eta_i + \mu_t + \varepsilon_{it}, \qquad (1)$$

with $i = 1, \dots, N$ denoting countries and $t = 1, \dots, T$ representing years, and where

- $INEQ_{it}$ is a measure of inequality (Gini and P90/P50) of household welfare within country i at a certain point in time t.
- $\Gamma_{it} \times Inc\_Lev_{it}$ is the categorical variable that indicates if inequality in that country i and that year t is measured with income or consumption interacted with the four income levels (lower, lower middle, upper middle, higher).
- X_{it} is the vector of population characteristics (population size and age dependency).
- Y_{it} is the vector of economic variables (GDP per capita growth, trade and trade composition, structure of the economy).
- REG_i indicates the regional classification of each country.
- η_i and μ_t stand for the inclusion of country and time dummies, respectively. (These occasionally entail, as fixed effects, a large variety of country-specific attributes and year-specific effects.)
- ε_{it} represents the error term.

For both inequality indicators, we first use ordinary least square (OLS) regression with pooled cross sections. However, since simple pooled OLS approaches have been judged unsatisfactory by many authors of multicountry studies, we also produce estimates using a panel random-effects model; we run it first on the whole sample, and then we restrict the analysis (1) to countries below the high-income threshold and (2) to observations after 2000. We opted for random effects in order to take into account the variation between countries and the effect of factors such as institutions, which are constant over time but differ between countries (Nielsen and Alderson 1995; Alderson and Nielsen 2002).

Table 2 reports the results of the different regressions (pooled OLS and panel random effects) for both the Gini and the P90/P50 ratio. To capture the specific effect of using consumption in SSA, the interacted dummies $\Gamma_{it} \times Inc\_Lev_{it}$ for consumption are further subdivided into the group of SSA countries and the group of all the others. For the interacted dummies, the baseline is represented by countries with low gross national income (GNI) using income; in this group, the vast majority are highly inegalitarian Central and Latin American countries. As mentioned, our two main purposes are to show that (1) all

other things been equal, using consumption underestimates SSA's countries inequality, and (2) this underestimation is more accentuated when considering the upper tail of the welfare distribution.

Overall, both hypotheses are verified. Irrespective of the model or the sample (including or excluding developed countries and restricting the observations to the post-2000 years), there is a clear SSA consumption effect. After controlling for country-level, year, regional and development-stage effects, inequality in SSA measured with consumption is always significantly lower than in other countries. For example, compared to the baseline, SSA's low-GNI countries using consumption record a 16 percent shortfall in Gini (table 2, column 2). Furthermore, SSA's low-GNI countries appear significantly less unequal than their Latin American peers (the baseline) but also less than the other low-GNI countries using consumption: these show a negative but smaller coefficient than those of SSA. When we restrict the sample to developing countries only or to developing countries after 2000,[11] the magnitudes of the SSA coefficients—all significant—further increase: between 20 and 30 percent depending on the model (table 2, columns 3 and 4).

Since we control for multiple effects, the gap looks more like a feature of how inequality is measured in SSA rather than a real *egalitarian effect* common to most of SSA's countries. This result is robust to different model specifications and also to changes in the dependent variable. When we change the dependent variable from the Gini to the P90/P50 ratio, the gap between SSA's consumption dummies and the others further increases; this confirms, that the gap between the two welfare measures increases if we consider a more (upper) tail sensitive measure.

When looking at the different random effects' models, all the estimated coefficients seem to have the expected signs. Growth is positive but insignificant unless we restrict the sample to developing countries after 2000, when it becomes positive and highly significant; most likely this captures the effect of globalization, when some developing countries such as China and India posted very high growth but also saw a surge in inequality (Ding and He 2018). Countries with a higher share of people of workforce age tend to be more unequal, while those in the developing world receiving higher percentages of remittances are more equitable (Azizi 2021). Finally, all other things being equal, countries with higher percentages of services over value added tend to be more unequal than others. This might capture the distributional

---

[11] The year 2000 is symbolically used as the starting year of the globalization period, as well as a year after which the number of observations substantially increases.

effect that Rodrik (2016b) defines as "premature deindustrialization." In traditional structural transformation, the growth of industry, especially manufacturing, played a big role. Several developing countries did not follow the development pattern of already industrialized countries and basically skipped the transition from agriculture to manufacturing, going straight to service sector development. The modern service sector in the developing world, however, is highly polarized, with few very well-paid jobs and a vast majority of low-paid jobs in retail and construction (Molini and Paci 2015; World bank 2019).

4 Our Proposal to Reduce the Inequality Underestimation

4.1 Introduction

The previously provided empirical estimations clearly demonstrate that the choice of the consumption or income variable is far from neutral. We propose a solution to overcoming the emerging potential mismesaurement. In the past decades several methods have been suggested to include the welfare of hard-to-survey populations, in particular, the extremely rich (e.g. Hlasny and Verme 2021; and references therein). The most famous method is to compare top incomes in household surveys with tax records (see, for example, Atkinson et al. 2011). In developing countries, where generally it is difficult to obtain this type of information from tax authorities, analysis on top incomes started later than in developed countries but has recently gained momentum (see, for example, Leigh and Van der Eng 2009; Alvaredo 2010; and Sanhueza and Mayer 2011).

More recently, Clementi et al. (2020) depart from the top-income literature and present recalibration exercises in order to involve a much bigger portion of the distribution. The principal novelty of this approach consists in the fact that, different than in other works, the reestimation of the top incomes tail is based on parameters coming from the same sample. In what follows, we first outline the methodology proposed by Clementi et al. (2020) for the purposes of analyzing inequality in SSA and then present the results of its applications to the sample of African countries considered in this study.

4.2 Multiple-Imputation Approach to Inequality Estimation and Inference

Schematically, the approach to inequality measurement proposed in this paper goes through the following steps.[12]

---

[12] The approach introduced by Clementi et al. (2020) is adapted from earlier work by Jenkins et al. (2011), who proposed a parametric multiple-imputation method to measure income inequality with right-censored (top-coded) data. This method has

First, by means of model selection techniques, the best-fitting parametric model for the consumption and income distributions of each country and year is selected. The models that are fitted to microdata belong to the family of generalized beta distributions introduced by McDonald and Xu (1995a, 1995b), which includes the four-parameter generalized beta II distribution (GB2) with probability density function:

$$f(x; a, b, p, q) = \frac{ax^{ap-1}}{b^{ap}B(p,q)[1+(x/b)^a]^{p+q}}, x > 0, \qquad (2)$$

and cumulative distribution function:

$$F(x; a, b, p, q) = I_z(p, q), z = (x/b)^a, x > 0, \qquad (3)$$

where $B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$ is the (complete) beta function, $\Gamma(\cdot)$ is the gamma function, and $I_z(p, q) = \frac{B(z;p,q)}{B(p,q)}$ is the regularized incomplete beta function—which is the ratio of the incomplete and complete beta functions. All four parameters are positive, with b being the scale parameter and a, p, and q being the shape parameters. The GB2 distribution is a flexible functional form incorporating many distributions as special cases. Of these, we focus on the three-parameter models of Singh and Maddala (1976) and Dagum (1977), which are often used in the income distribution literature and can be obtained as special cases of the GB2 for, respectively, $p = 1$ and $q = 1$.[13]

Second, once the best-fitting parametric model for both consumption and income data has been selected, the approach uses the model's parameter estimates to derive imputed values for observations above some lower-bound consumption threshold defining (in absolute terms) a minimum middle-class standard of

been shown to be effective in recovering the population income distribution accurately in any given year, but it originates from cross-sectional data applications and is not suited to be applied to longitudinal data (e.g. Tan 2021).

[13] For details, see McDonald (1984), McDonald and Xu (1995a, 1995b), Kleiber and Kotz (2003), and McDonald and Ransom (2008). Of particular importance in the current context, it is the desirable behavior of the GB2 and related distributions in their upper tail, which is heavy in that it decays like a power function as the size variable increases, rather than decaying exponentially fast ,like, for instance, the log-normal distribution with a mildly heavy upper tail. For more on this, see, for example, Kleiber (1996), Schluter and Trede (2002), Kleiber and Kotz (2003), and Kleiber (2008).

living.[14] For the purposes of this study, two absolute thresholds are used to define the middle class in SSA:[15]

- Per capita daily consumption greater than US$5.5 in 2011 PPP, which includes both the lower- and the upper-middle class

- Per capita daily consumption greater than US$10 in 2011 PPP, which identifies the upper-middle class

Imputed values for observations above these thresholds are derived by means of the so-called inverse transform method. That is, given the fitted model, the cumulative distribution function for each observation i above the consumption threshold t_c is, using standard notation for left-truncated distributions, as follows:

$$G(x_i; \hat{\theta}) = u_i = \frac{F(x_i; \hat{\theta}) - F(t_c; \hat{\theta})}{1 - F(t_c; \hat{\theta})}, x_i > t_c, u_i \in [0,1), \tag{4}$$

where $\hat{\theta} = \{\widehat{a_y}, \widehat{b_c}, \widehat{p_y}, \widehat{q_y}\}$ is the set of parameter estimates and the subscripts c and y refer to consumption and income, respectively.[16] Inverting, one gets:

$$G^{-1}(u_i; \hat{\theta}) = x_i = F^{-1}\{u_i[1 - F(t_c; \hat{\theta})] + F(t_c; \hat{\theta}); \hat{\theta}\}, x_i > t_c, u_i \in [0,1). \tag{5}$$

Thus, a value of x_i for each observation above the consumption threshold is generated by substituting into this expression a value of u_i that is equal to a random draw from a standard uniform distribution.[17]

---

[14] As already discussed in Clementi et al. (2020), opting for such a definition of the middle class in the context of developing countries seems reasonable for at least two orders of reasons First, unlike for developed countries, we cannot use relative welfare measures for defining the middle class, since in developing countries the latter does not often coincide with some function of the distribution's median (that is, the middle class does not generally occupy the center of the distribution). Scholars thus often opt for absolute measures. Second, a further complication one might encounter in developing countries is defining an upper bound. As already anticipated, these countries often focus their attention on getting consumption data right and disregard income data collection. Since consumption is very accurate in capturing the well-being of poorer people, while it is rather imprecise in capturing that of people living in the upper percentiles, it follows that when defining the middle class in these countries it seems reasonable to opt for a lower-bound threshold (rather than an interval) of the type "middle class and above" and leave the border between middle class and upper class somehow undefined.

[15] The reason for working with two thresholds is that using only the second one could lead to rather conservative estimates of inequality, as the correction of the consumption data in this case typically affects a tiny group of households at the far end of the distribution. Instead, by using also the first threshold, one can impute income variability to the original data for a broader group of households, which prevents a potentially downward-biased estimation of inequality.

[16] Clearly, $p_y = 1$ for the Singh-Maddala distribution and $q_y = 1$ for the Dagum. The Clementi et al.'s (2020) approach is designed to alter the *shape* of the consumption distribution at the top end, but not its scale. That is why the set of parameter estimates used for imputing values above the consumption thresholds includes the shape parameter estimates for the income distribution of each country and year ($\widehat{a_y}$, $\widehat{p_y}$ e $\widehat{q_y}$) and the estimated scale parameter for the consumption distribution ($\widehat{b_c}$).

[17] In equations (4) and (5), the values of the GB2 cumulative distribution function at the truncation point t_c, $F(t_c; \hat{\theta})$ and those for each x above the consumption threshold, $F(x_i; \hat{\theta})$ are estimated by inserting parameter estimates into equation (3). The cumulative distribution functions in the cases of the Singh-Maddala and Dagum distributions are given by simpler expressions and can be found, for instance, in Kleiber and Kotz (2003, ch. 6).

Third, the imputed values for observations above the consumption thresholds are combined with observed expenditures for those lying below to produce partially synthetic datasets for each country and year to which complete-data methods can be applied for estimating inequality statistics, such as the Gini coefficient, the MLD, and the Theil index.

Finally, numerous repetitions of the second and third steps (to control for the randomness of each partially imputed data set) produces R synthetic data sets for each country-year pair and, correspondingly, R sets of inequality estimates that can be combined using the following rule (Reiter 2003; An and Little 2007):

$$\bar{q}_R = \frac{1}{R}\Sigma_{r=1}^R \hat{q}_r, \tag{6}$$

which is the simple average of the point estimates $\hat{q}_r$ that are derived using complete-data methods from each of the R partially synthetic datasets. Furthermore, the variance of $\bar{q}_R$ is estimated using the following:

$$\text{var}(\bar{q}_R) = b_R/R + \bar{v}_R, \tag{7}$$

where:

$$\bar{v}_R = \frac{\Sigma_{j=1}^R v_j}{R} \tag{8}$$

is the average of the sampling variances and

$$b_R = \frac{\Sigma_{j=1}^R(\hat{q}_j - \bar{q}_R)^2}{R-1} \tag{9}$$

is additional variability reflecting the finite number of imputations R (for example, Reiter 2003, 5; An and Little 2007, 926; Jenkins et al. 2011, 71).[18] In the next section, we report estimates of inequality indices derived from the imputation-augmented data based on $R = 1,000$ repetitions.

4.3   Estimates of SSA Inequality, 2004−17

This section is divided in two parts. In the first, we look at whether the proposed parametric distributions fit the original survey data for both income and consumption. In the second, we calculate a set of distributional indicators on the original consumption data and compare them to the corrected ones.

---

[18] Hence, the square root of (7) is an estimate for the standard error of $\bar{q}_R$.

4.3.1 Parameter estimation, model selection, and goodness of fit

All generalized beta models considered in this paper were fitted to consumption and income distributions using maximum likelihood estimation. For fitting models to data, we have used Stata's programs developed by Jenkins (1999, 2007, 2014). These programs maximize the loglikelihood numerically and estimate parameter variance using the negative inverse Hessian. A number of distributional measures implied by fitted models were also obtained using the Stata's commands developed by the author.

Tables 3 and 4 present our estimates of models' parameters together with their standard errors, the values of loglikelihood ($\ln L$) at last iteration, and model selection criteria such as the Akaike (1973) and Bayesian (Schwarz 1978) information criteria (AIC and BIC, respectively).[19] In order to compare the fit of the GB2 model and its nested alternatives (the Dagum and Singh-Maddala), we also give the results of likelihood ratio tests for the fitted models. The likelihood ratio statistics takes the form:

$$2\big[\mathrm{ln}L\big(\hat{\theta}_U\big) - \mathrm{ln}L\big(\hat{\theta}_R\big)\big] \sim \chi^2(h), \tag{10}$$

where $\ln L\big(\hat{\theta}_U\big)$ and $\ln L\big(\hat{\theta}_R\big)$ are, respectively, the log-likelihood values corresponding to the unconstrained (GB2) and nested or restricted models (Dagum and Singh-Maddala), $\hat{\theta}$ is the set of estimated parameters, and h is the difference in the number of parameters in the two compared models (equal to 1 in our setting). The differences between GB2 and its nested alternatives can be thus compared using a chi-square (χ^2) distribution with one degree of freedom. In the tables, asterisks are placed next to the log-likelihood values of the Dagum and/or the Singh-Maddala distribution if the improvement gained in adding a further parameter is of practical significance at the 5 percent level—that is, if the GB2, with its fourth parameter, provides a statistically significant (at the 5 percent level) better fit over the Dagum and/or the Singh-Maddala distribution.[20]

The results of model selection for consumption distributions, presented in table 3, suggest that the GB2 model is a better fit to survey data in all countries and years except for Burkina Faso in 2014, Malawi in 2013 and 2017, and Niger in 2011, where the Dagum model is as good as the GB2, while for the 2014

---

[19] The expressions for the loglikelihood of the GB2 and its nested models (the Singh-Maddala and Dagum) are given in Kleiber and Kotz (2003). Model selection criteria will select, when comparing models with the same number of parameters, the model with the smallest $l = -\ln L$ according to the formula $(2 \times l) + (d \times k)$, where k represents the number of parameters in the fitted model and $d = 2$ for the usual AIC or $d = \ln L$ (N being the number of observations) for the so-called BIC. Hence, when comparing models fitted by maximum likelihood to the same data, the smaller the AIC or BIC the better the fit.

[20] The critical value of the $\chi^2(1)$ distribution is 3.84 at the 5 percent level.

Ethiopian data the Singh-Maddala ranks first (being observationally equivalent to the GB2). For income data, the results in table 4 are somewhat mixed. The GB2 is clearly the best model for Burkina Faso, Ethiopia, Ghana, Malawi (for 2017 income data), Nigeria, Rwanda, Tanzania, and Uganda, whereas for Kenya, the Malawi's 2013 household incomes and the Niger's 2011 data, the Singh-Maddala seems to be as good as the GB2. A similar conclusion applies to Niger's 2014 data, but in this case the Dagum model fits the data better than the alternatives. In general, the GB2 model gives the best fit to both consumption and income data in 7 of the 14 cases analyzed: Ethiopia 2014, Ghana 2005, Nigeria 2004, Rwanda 2014, Tanzania 2009, Tanzania 2013, and Uganda 2005. For the remaining country-year pairs, the following parametric distrbutions are instead selected as imputation models:

- Burkina Faso 2014: Dagum for the consumption distribution and GB2 for the income distribution
- Ethiopia 2014: Singh-Maddala for the consumption distribution and GB2 for the income distribution
- Kenya 2005: GB2 for the consumption distribution and Singh-Maddala for the income distribution
- Malawi 2013: Dagum for the consumption distribution and Singh-Maddala for the income distribution
- Malawi 2017: Dagum for the consumption distribution and GB2 for the income distribution
- Niger 2011: Dagum for the consumption distribution and Singh-Maddala for the income distribution
- Niger 2014: GB2 for the consumption distribution and Dagum for the income distribution

The suitability of fit of the functional forms chosen according to the model selection methods is evaluated by comparing the sample values of distributional indicators reported in table 1 with their counterparts implied by the fitted models—see the last four columns of table 3 and table 4.[21] Specifically, in figures 3 and 4 the comparison relies on checking for overlap between 95 percent confidence intervals of theoretical and sample indicators to draw conclusions about the accuracy of selected distributional statistics deduced by parameter estimates. The results suggest that for most of the indices (the mean, the Gini coefficient, the MLD, and the Theil index), the best-fitting models produce theoretical values that are quite often in a close agreement with the corresponding sample values—the respective confidence intervals overlap in a way that let us exclude that the predicted values, and the sample estimates of chosen

---

[21] The analytical expressions for all indices considered here, which are functions of the estimated parameters of the GB2 and its nested distributions, can be found, inter alia, in Kleiber and Kotz (2003, ch. 6) and Jenkins (2009).

indicators can be considered different. The most notable exceptions are the theoretical estimates implied by the best-fitting GB2 model for 2017 Malawian and 2005 Ugandan incomes, which differ significantly from the corresponding sample estimates. As shown by the quantile-quantile plots in figure 5, this fact could reflect the poor performance of the GB2 at the top of the Malawian and Ugandan income distributions, where there is a systematic departure of empirical observations from the theoretical predictions of the assumed specification. However, the results for the nested three-parameter Singh-Maddala and Dagum distributions (not shown here but available on request from the authors) are even worse, especially for higher quantiles. This explains why we shall keep using the GB2 distribution for imputing observations in the top part of the 2017 Malawian and 2005 Ugandan welfare distributions.

4.3.2 Multiply-imputed estimates of inequality

Table 5 and figure 6 display the simulation results by country using the middle-class thresholds set at US$5.5 and US$10 per day in 2011 PPP terms. As discussed in section 4.1, with the US$10 threshold, the correction of consumption data applies to a smaller group of households than with the US$5.5 threshold (compare shares in columns 5 and 12 of table 5). The estimated inequality using the US$5.5 threshold is clearly higher, since more information is taken from the income distribution and, as discussed before, income tends to have higher variability than consumption.

For example, in Burkina Faso, where according to the US$5.5 threshold the middle-class group would account for about 8 percent of the population, the correction of consumption for this group would lead to a Gini of 0.47 from 0.36 in the original consumption data (compare column 5 of table 1 and column 6 of table 5). On the other hand, when using the US$10 threshold, only 2 percent of the households will see their consumption corrected, and the obtained Gini is 0.42 (column 13 in table 5). Likewise, in all analyzed countries, the two thresholds define an upper and lower bound for the simulated Gini, where the upper bound is obtained from the US$5.5 threshold and the lower is obtained from the US$10 threshold (figure 6).

Figure 7 and figure 8 display the impact of the correction on the original consumption data. In figure 7, the correction is applied on the middle-class group defined by the US$5.5 threshold, whereas in figure 8 it is applied using the US$10 threshold. Correcting consumption implies using from the middle-class thresholds onward (vertical dashed line) the parametrized tail derived from the corresponding income distribution (gray squares), which in all the analyzed surveys lays above that of the consumption distribution. The difference is very clear when cutting the consumption distribution at US$5.5 (figure 7),

but less pronounced when using the US$10 threshold (figure 8). As a consequence, when using the US$5.5 threshold, we reestimate a bigger chunk of the distribution and introduce in this way more variability than in the case of US$10, leading to a bigger increase of the Gini and other inequality measures (see table 5).

Two points are important to highlight: the magnitude of the inequality measurement correction and what the comparison over time can tell us. Regarding the first, in section 3 we calculated that the inequality underestimation caused by using consumption rather than income was in the order of 15 or 16 percent: the Gini of a low-income SSA country using consumption is 16 percent lower than what the model would predict for a country with similar characteristics but using income, 15 percent in case of a low-middle-income SSA country using consumption. If we compare the "corrected" Gini's from table 5 with the original ones in table 1, we observe that with a threshold of US$5.5 the average variation upward is of about 16 percent, while with a US$10 threshold the variation upward is of about 10 percent. Our sample of SSA countries is clearly smaller than that used for the regressions in table 2, but it is very reassuring that the correction comes very close to compensating for the predicted inequality underestimation.

Regarding the second point—what the comparison over time can tell us—the particular nature of underestimation (in upper tails consumption biased downward) makes inequality increase as economic growth leads to growth in either the size or in the economic power of the groups in the upper tail (above the middle-class lines). Specifically, as more people are pushed into the middle-class segment of the distribution, they increasingly enter an area of the distribution where, we argue, consumption tends to underestimate their welfare. Therefore, the very process that drives an increasing disparity between those at opposite poles of the consumption distribution can also compromise the ability of consumption to capture the extent of these disparities. The comparison of the four countries where we have two survey rounds (Ethiopia, Malawi, Niger, and Tanzania) confirms that this effect is present. All countries posted fast GDP growth (around 6 percent) and inequality measured on consumption increases in Malawi only, but when corrected, inequality increases over time in all countries except Ethiopia. However, when corrected, the Gini value of Ethiopia in 2016 is insignificantly different from that in 2014, but when measured on consumption, it showed a 5 percentage points decline. While at present we provide some initial evidence of this problem and a method to attenuate the issue of cross-sectional inequality underestimation, we leave it for future research to understand how this measurement issue interacts with

distributional changes—that is, how distributional changes that accompany growth interact with this measurement issue in a dynamic sense.

5   Discussion and conclusions

In summary, we have laid out a puzzle in the relationship among growth, inequality, and poverty in SSA and suggested that an inequality measurement problem lies at its heart. Consumption-based inequality measures miss important information at the top end of the welfare distribution, leading to an underestimation of inequality. Besides identifying the problem, we propose an economically viable solution to reduce this underestimation; we suggest that using information from income distributions collected under the FAO-RuLIS project to recalibrate the top end of the consumption distribution may provide a practical solution to this measurement problem. Overall, we calculate that by using consumption as a welfare measure, inequality is biased downward by 16 percent in SSA when using Gini as the preferred inequality measure and by about 20 percent when using the 90th percentile over the 50th percentile ratio. When correcting the inequality measures with the proposed method, we almost eliminate this downward bias. The evidence is certainly not conclusive, since we can apply the correction to only a small set of countries, but it sheds new light on the distributional changes in SSA in the past two decades. The value added of our work compared to similar contributions (among others, Chancel et al. 2019) lies in the fact that we use the original distributions as much as possible (consumption but also income), only in-sample information, and limit the amount of theoretical assumptions by "letting the data speak."

We conclude, however, on a more speculative note, posing some open questions that we hope will contribute to framing a research agenda centered on the revaluation of distributional issues in light of our new findings. What appears on the surface as a somewhat technical measurement issue has, in fact, implications that go far deeper. In this paper we have argued that the standard empirical toolkit available to development economists working on SSA has limited our ability to appreciate the magnitude and persistence of inequalities in the continent. We hope that, by beginning a process of refining and expanding this toolkit, we will have helped put in motion a process that will overcome what we believe is a technical bottleneck to understanding the effects of inequality in SSA.

Refocusing attention on inequality in SSA can have an effect in both academic and policy spaces. In the world of academic research, we hope to see more attention being placed on the collection and analysis

of data that can illuminate the nature, evolution, and consequences of inequality in SSA. However, while it will be important to improve our understanding of inequality in Africa, we also believe that we already have enough evidence to be confident that inequality is playing a key role in the persistence and reproduction of poverty in SSA. With this in mind, researchers will need to dedicate attention to the distributional patterns of the growth process itself, and to ways to increase the inclusivity of this process.

In addition to issues of distribution and growth, the research agenda that emerges will also need to dedicate attention to understanding the scope for post-outcome redistribution. This research agenda will be of interest to those in the policy sphere. The issue of taxation as a means of resource redistribution, in particular, will need to be informed by a clearer understanding of the scope for expanding the fiscal space through progressive taxation. This, in turn, will require more evidence than that which we currently have on the wealth held and income captured by top earners, much of which currently goes untaxed (Alstadsætera et al. 2014). Ultimately, a focus on the nexus between distribution and poverty may illuminate what potential Africa has for endogenous poverty alleviation, rather than a reliance on aid that, to the extent that it has led to poverty alleviation, has largely done so without a concomitant structural transformation necessary for sustained poverty reduction and inclusive growth.

Figures

Figure 1: Income and Consumption by Levels of Gross National Income (GNI) per Capita (Atlas method)

Source: World Bank.

Figure 2: Differences between Inequality Measures of Income and Consumption over Same-Year Surveys (%)
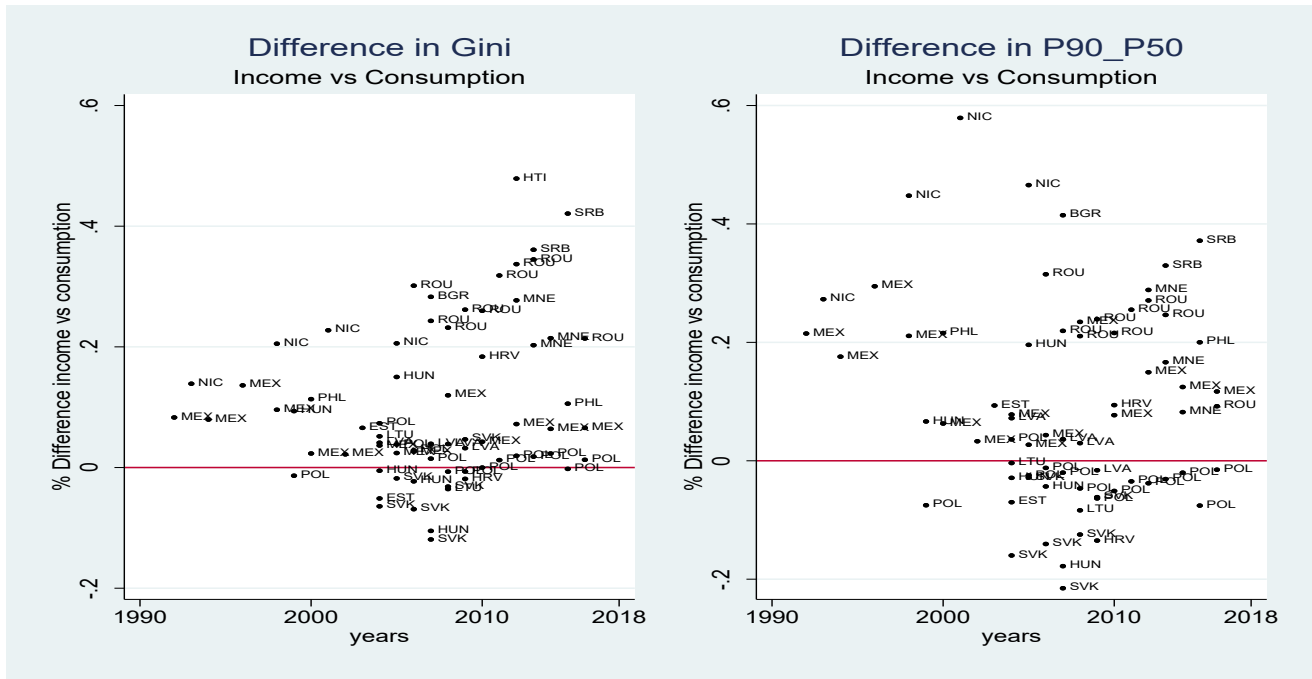
Source: World Bank.

Figure 3: Comparison between the Sample Values of Chosen Distributional Indicators and Their Counterparts Implied by the Best-Fitting Consumption Models
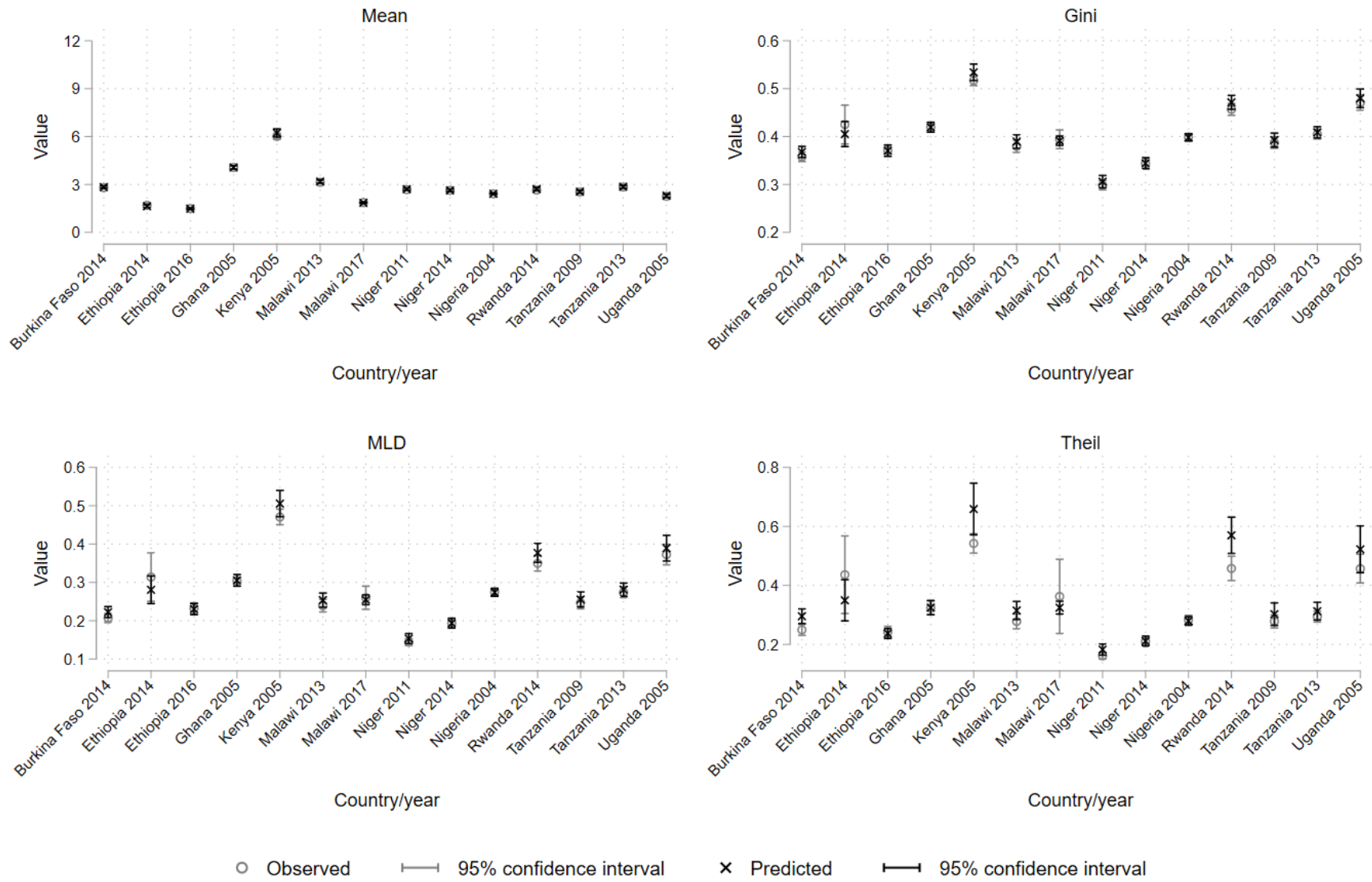
Source: World Bank.

Figure 4: Comparison between the Sample Values of Chosen Distributional Indicators and Their Counterparts Implied by the Best-Fitting Income Models
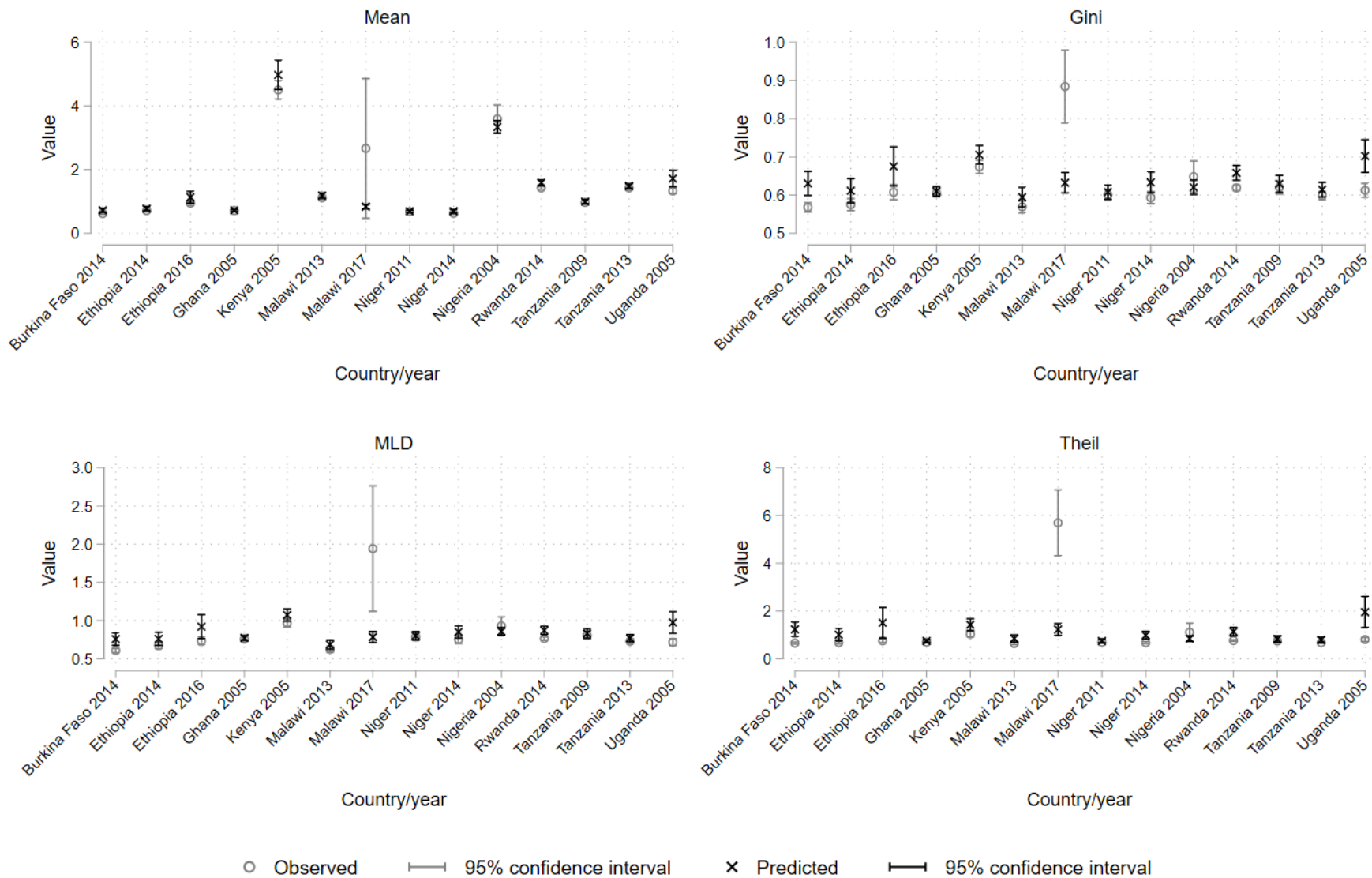
Figure 5: Quantiles of per Capita Daily Income (in 2011 PPP) against the Quantiles of the GB2 Distribution for Malawi 2017 and Uganda 2005
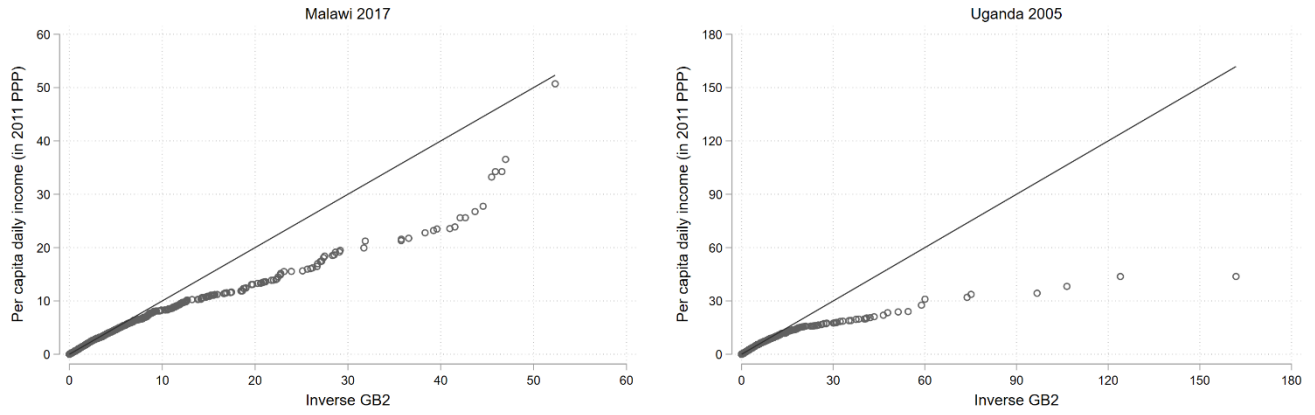
Source: World Bank.
Note: PPP = purchasing power parity; GB2 = generalized beta II.

Figure 6: Inequality Estimates Derived from Repetition of the Imputation Process R = 1,000 Times, by Index and Country

a. Gini

b. MLD

c. Theil
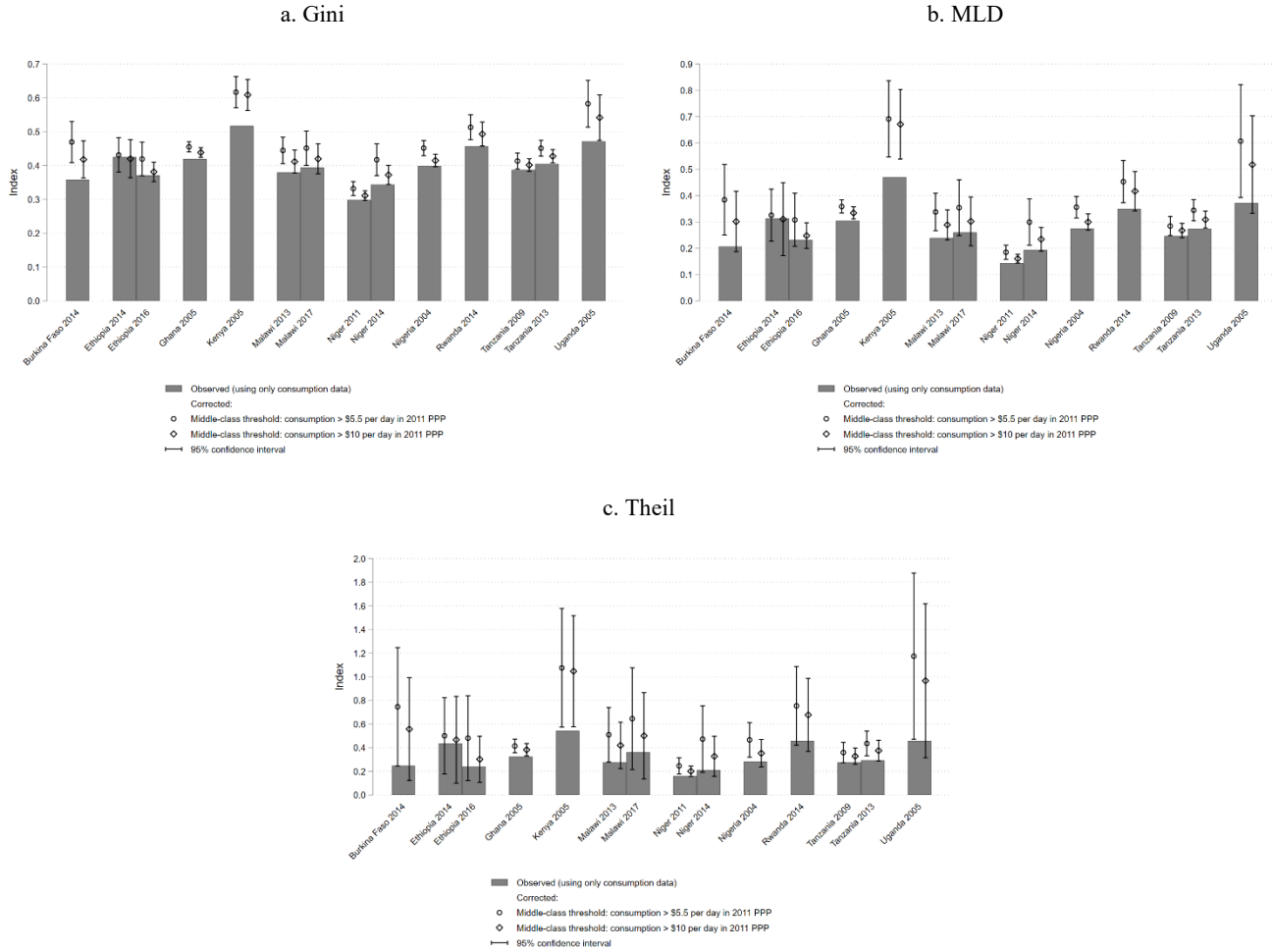
Source: World Bank.

Note: The associated 95 percent confidence intervals are obtained using a normal approximation, given the very large number of both sample sizes and imputations R. MLD = mean log deviation.

Figure 7: Partially Synthetic Welfare Distributions Based on R = 1,000 Repetitions of the Imputation Process (US$5.5 per day threshold)

a. Burkina Faso 2014

b. Ethiopia 2014

c. Ethiopia 2016

d. Ghana 2005

e. Kenya 2005

f. Malawi 2013

g. Malawi 2017

h. Niger 2011

i. Niger 2014

Figure 7: Continued

j. Nigeria 2004
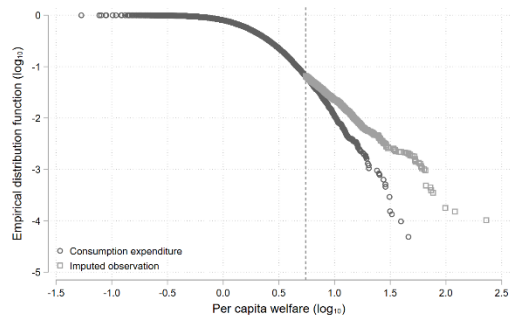
k. Rwanda 2014
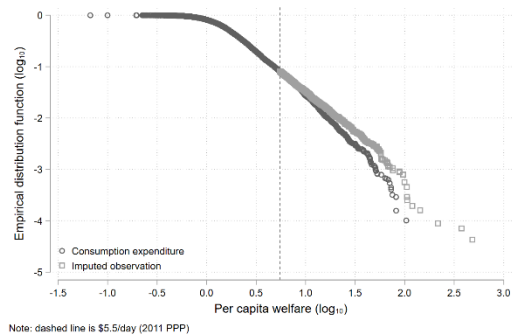
l. Tanzania 2009
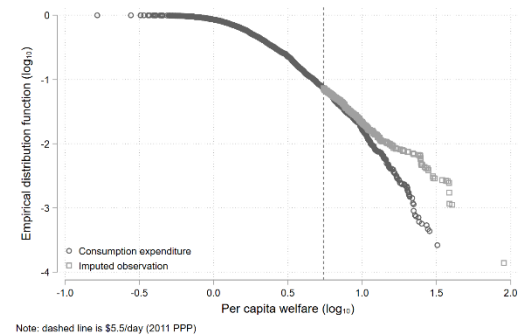
m. Tanzania 2013
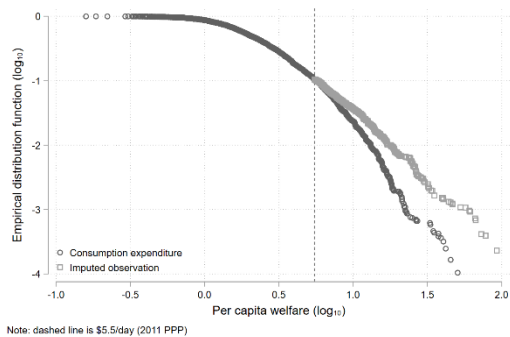
n. Uganda 2005
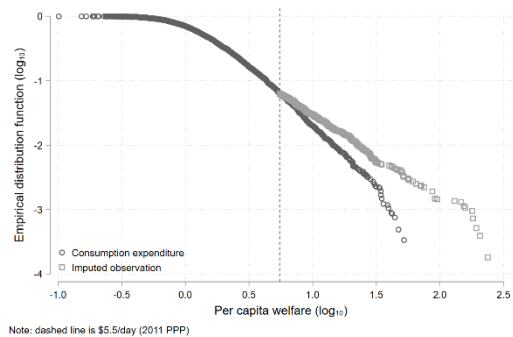
Source: World Bank.

Figure 8: Partially Synthetic Welfare Distributions Based on R = 1,000 Repetitions of the Imputation Process (US$10 per day threshold)

a. Burkina Faso 2014

b. Ethiopia 2014

c. Ethiopia 2016

d. Ghana 2005

e. Kenya 2005

f. Malawi 2013

g. Malawi 2017

h. Niger 2011

i. Niger 2014
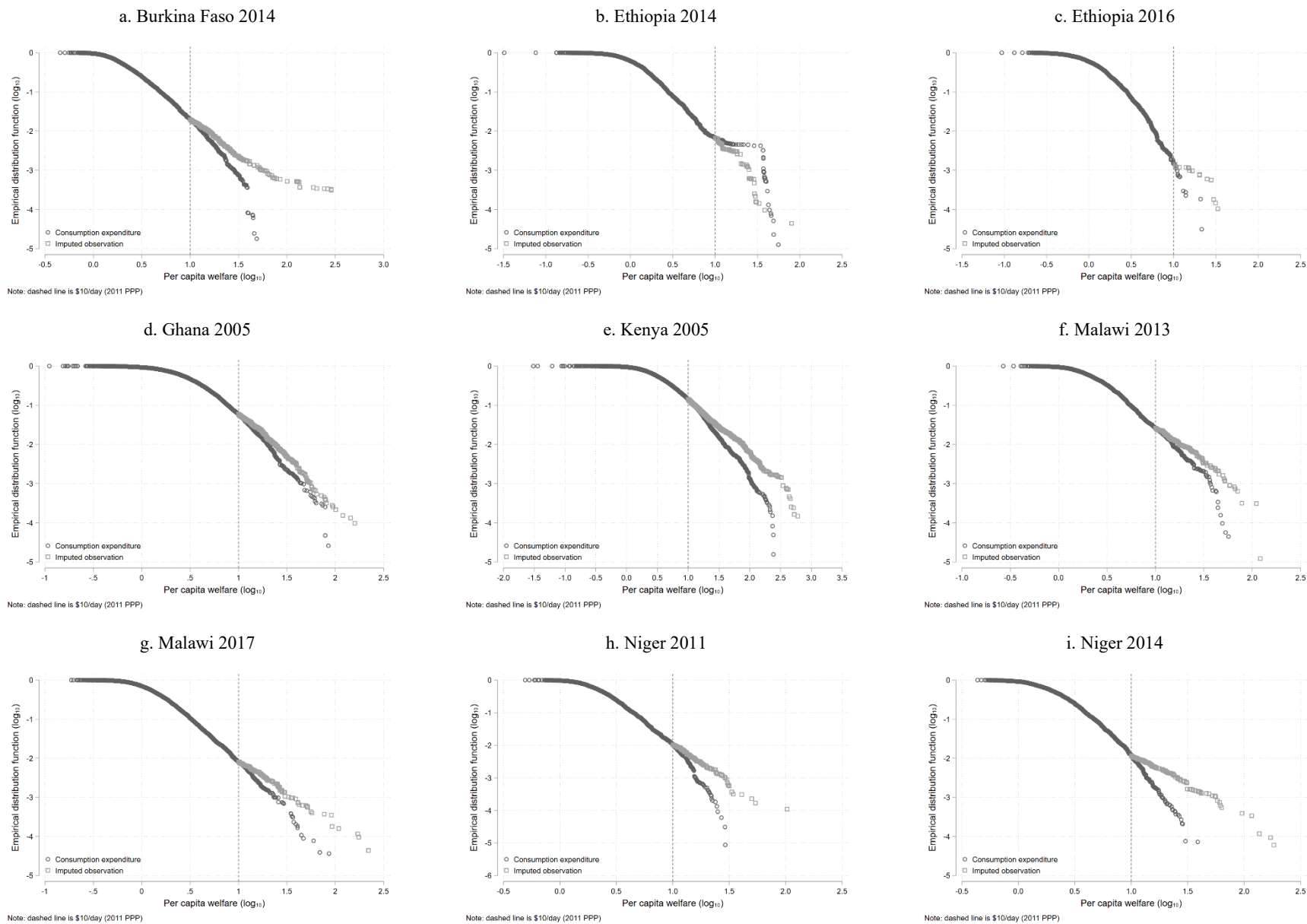
Figure 8: Continued

j. Nigeria 2004
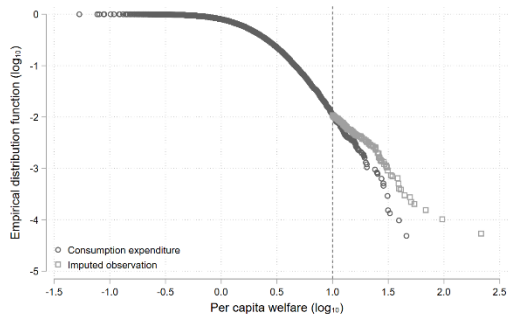
Note: dashed line is $10/day (2011 PPP)

k. Rwanda 2014
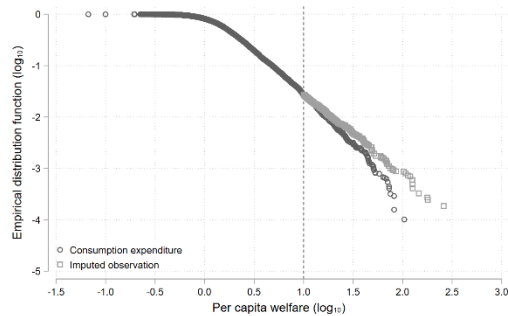
Note: dashed line is $10/day (2011 PPP)

l. Tanzania 2009
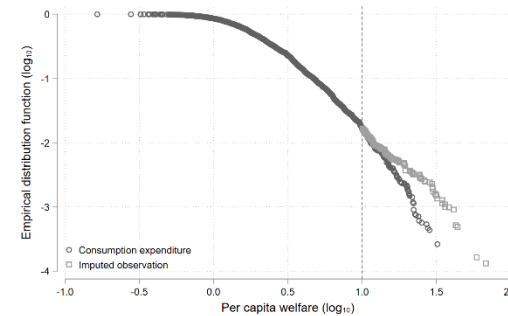
Note: dashed line is $10/day (2011 PPP)

m. Tanzania 2013
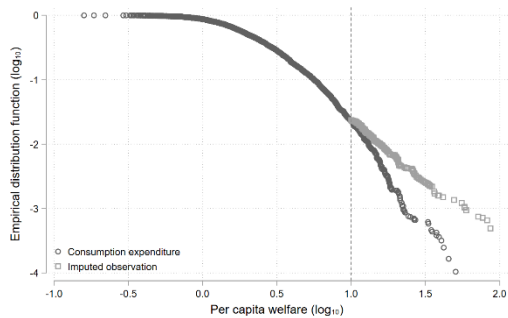
Note: dashed line is $10/day (2011 PPP)

n. Uganda 2005
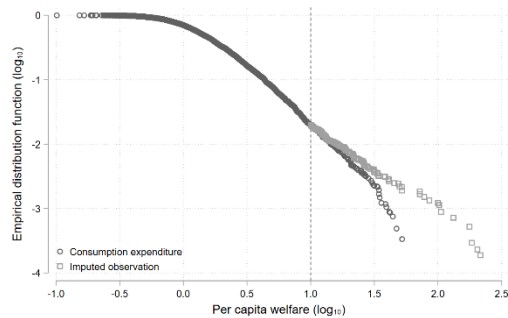
Note: dashed line is $10/day (2011 PPP)

Source: World Bank.

Tables

Table 1: Distributional Summary Statistics for the Survey-Based Consumption and Income Variables Used in the Analysis

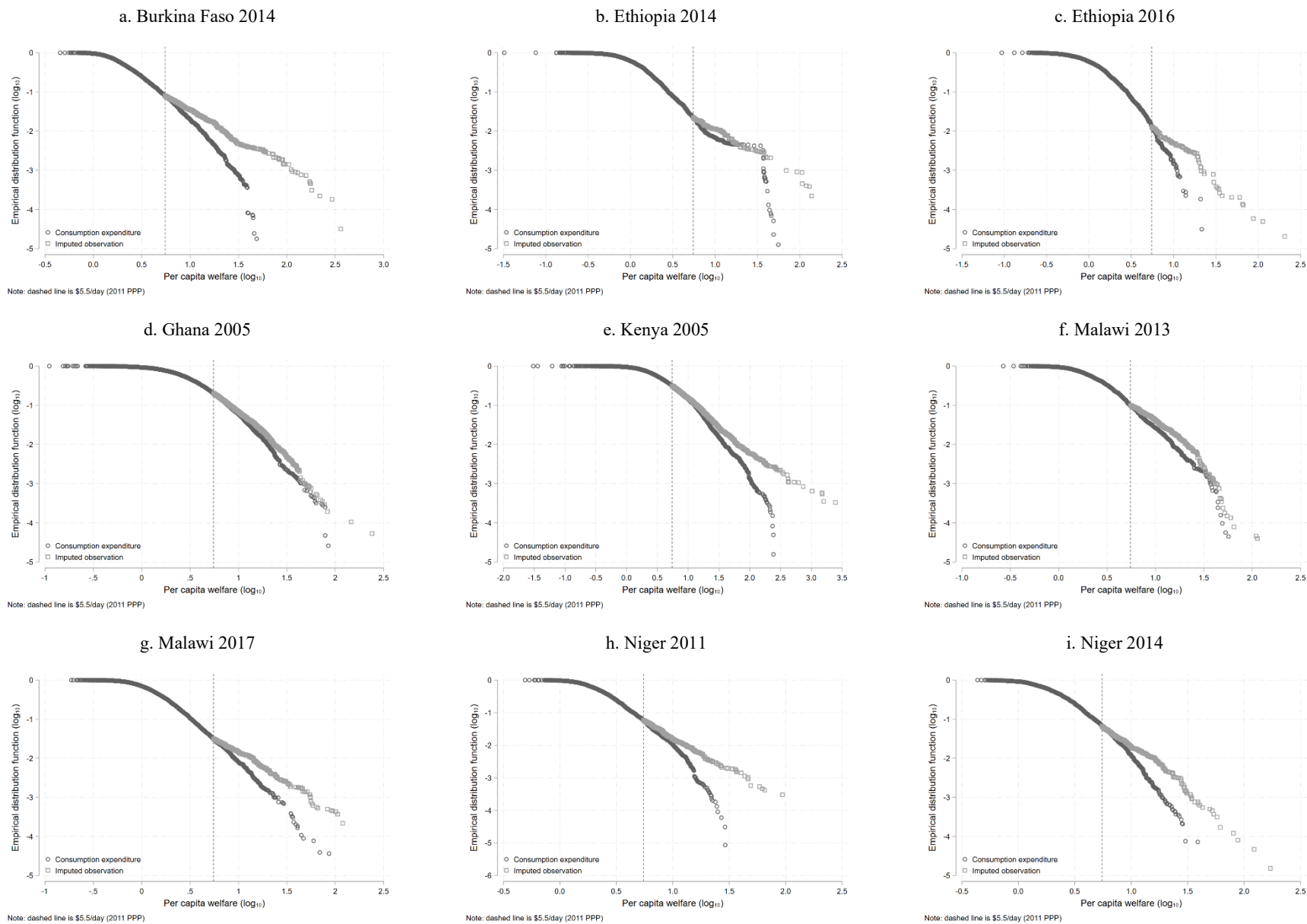| Country	Year	Households[a]	Consumption				Income			
			Mean[b]	Gini	MLD	Theil	Mean[b]	Gini	MLD	Theil
Burkina Faso	2014	9,229	2.796	0.358	0.207	0.249	0.620	0.568	0.608	0.658
Ethiopia	2014	4,733	1.694	0.425	0.314	0.436	0.712	0.575	0.674	0.675
Ethiopia	2016	4,262	1.489	0.371	0.232	0.242	0.943	0.607	0.736	0.754
Ghana	2005	7,659	4.067	0.419	0.305	0.324	0.712	0.605	0.759	0.698
Kenya	2005	11,700	6.009	0.517	0.471	0.542	4.500	0.674	0.971	1.037
Malawi	2013	3,873	3.135	0.380	0.239	0.278	1.115	0.569	0.625	0.642
Malawi	2017	11,783	1.864	0.394	0.260	0.363	2.665	0.884	1.941	5.688
Niger	2011	3,538	2.679	0.299	0.144	0.161	0.680	0.601	0.789	0.692
Niger	2014	3,290	2.632	0.344	0.194	0.211	0.627	0.593	0.753	0.672
Nigeria	2004	16,922	2.419	0.399	0.275	0.283	3.593	0.648	0.932	1.113
Rwanda	2014	14,174	2.646	0.457	0.350	0.458	1.426	0.619	0.765	0.762
Tanzania	2009	3,037	2.529	0.387	0.247	0.278	0.969	0.620	0.808	0.745
Tanzania	2013	4,587	2.840	0.405	0.274	0.294	1.426	0.599	0.729	0.674
Uganda	2005	7,199	2.257	0.471	0.373	0.457	1.328	0.612	0.717	0.807

Source: World Bank, based on microdata from RIGA/RuLIS database.
Note: MLD = mean log deviation.
a. Effective number of observations after removal of missing and nonpositive values.
b. Dollar-a-day money amount in 2011 purchasing power parity.

Table 2: Pooled OLS and Panel Random-Effects (RE) Regressions for the Gini and the P90/P50 Ratio

Independent variable	Gini				P90/P50		
	Full model, OLS	Full model, RE	No high income, RE	No high income and post-2000, RE	Full model, RE	No high income, RE	No high income and post-2000, RE
Low income with consumption (SSA)	-0.187***	-0.156***	-0.201***	-0.314***	-0.631***	-0.818***	-1.206***
Low-middle income with consumption (SSA)	-0.155***	-0.145***	-0.190***	-0.295***	-0.618***	-0.801***	-1.182***
Upper-middle income with consumption (SSA)	-0.219***	-0.151***	-0.225***	-0.317***	-0.562***	-0.877***	-1.206***
Low income with consumption (OTH)	-0.045***	-0.049***	-0.047***	-0.119***	-0.137***	-0.134***	-0.345***
Low-middle income with consumption (OTH)	-0.059***	-0.063***	-0.058***	-0.124***	-0.151***	-0.143***	-0.342***
Upper-middle income with consumption (OTH)	-0.064***	-0.077***	-0.072***	-0.135***	-0.167***	-0.157***	-0.354***
High income with consumption (OTH)	-0.049***	-0.065***	—	—	-0.171***	—	—
Low income with income	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline	Baseline
Low-middle income with income	-0.009	-0.020***	-0.020**	-0.064***	-0.060***	-0.066***	-0.221***
Upper-middle income with income	-0.043***	-0.041***	-0.040***	-0.099***	-0.135***	-0.140***	-0.341***
High income with income	-0.099***	-0.073***	—	—	-0.197***	—	—
Population in year	0.122***	-0.084	-0.103*	0.052		-0.244	0.242
Age dependency ratio, workforce over total population	0.075***	0.117***	0.141***	0.155***	0.339***	0.363***	0.468***
GDP per capita growth	0.019	0.028	0.041	0.077***	0.060	0.099	0.176**
Fuel exports, percentage over total merchandise	0.871	-0.117	0.493	2.111	-2.609	-1.221	4.358
Remittances, percentage over GDP	-0.113***	-0.098***	-0.095***	-0.034	-0.204**	-0.216**	-0.106

Table 2: Continued

Independent variable	Gini				P90/P50		
	Full model, OLS	Full model, RE	No high income, RE	No high income and post-2000, RE	Full model, RE	No high income, RE	No high income and post-2000, RE
Agriculture value added, percentage over total GDP	-0.025	-0.067**	-0.047	0.044	-0.193**	-0.145	0.002
Service value added, percentage over GDP	0.211***	0.134***	0.162***	0.123***	0.290***	0.368***	0.283***
Export of good and services, percentage over GDP	-1.512***	1.584**	3.934***	5.003***	-0.095	2.844	8.462***
region = EAP	0.060***	0.062***	-0.084***	-0.090***	0.088**	-0.191***	-0.199***
region = ECA	0.011**	0.023**	-0.120***	-0.116***	0.042	-0.252***	-0.227***
region = LAC	0.132***	0.151***	—	—	0.295***	—	—
region = MENA	0.026***	0.037**	-0.112***	-0.114***	0.042	-0.255***	-0.257***
region = SAS	-0.007	0.035*	-0.108***	-0.100***	0.064	-0.220***	-0.197***
region = SSA	0.216***	0.195***	0.095***	0.117***	0.650***	0.545***	0.672***
Constant	0.243***	0.242***	0.342***	0.395***	0.163***	0.379***	0.490***
Number of observations	1,529	1,529	1,026	741	1,529	1,026	741
R square							

Source: World Bank.

Note: OLS = ; RE = ; EAP = East Asia and Pacific; ECA = Europe and Central Asia, LAC = Latin America and the Caribbean; MENA = Middle East and North Africa, SAS = South Asia, SSA = Sub-Saharan Africa, OTH = other countries.

$*$ p<0.05, $**$ p<0.01, $***$ p<0.001.

Table 3: Maximum Likelihood Estimation of Generalized Beta Models for Consumption Distributions

Country	Year	Model[a]	$\widehat{a_c}$		$\widehat{b_c}$		$\widehat{p_c}$		$\widehat{q_c}$		ln L	AIC	BIC	Mean	Gini	MLD	Theil
Burkina Faso	2014	GB2	2.341	(0.350)	0.812	(0.230)	6.369	(3.621)	0.964	(0.183)	-15,186.9	30,381.7	30,410.2	2.844	0.369	0.224	0.299
		D	2.277	(0.042)	0.775	(0.091)	6.998	(1.529)	—		-15,186.9	30,379.8	30,401.2	2.839	0.368	0.223	0.296
		SM	5.190	(0.167)	1.473	(0.024)	—		0.367	(0.017)	-15,234.8*	30,475.6	30,496.9	2.969	0.399	0.269	0.409
Ethiopia	2014	GB2	2.830	(0.382)	1.031	(0.054)	1.065	(0.207)	0.788	(0.153)	-5,893.6	11,795.3	11,821.1	1.636	0.404	0.279	0.345
		D	2.391	(0.085)	1.011	(0.063)	1.366	(0.129)	—		-5,895.5	11,797.1	11,816.5	1.615	0.397	0.268	0.319
		SM	2.945	(0.110)	1.041	(0.049)	—		0.749	(0.063)	-5,893.7	11,793.5	11,812.9	1.639	0.405	0.281	0.350
Ethiopia	2016	GB2	0.467	(0.537)	0.238	(0.956)	32.299	(94.176)	15.555	(29.732)	-5,041.0	10,090.0	10,115.5	1.486	0.371	0.230	0.237
		D	2.460	(0.082)	1.043	(0.065)	1.222	(0.111)	—		-5,069.6*	10,145.2	10,164.2	1.542	0.392	0.263	0.306
		SM	2.635	(0.088)	1.169	(0.059)	—		0.990	(0.076)	-5,074.1*	10,154.1	10,173.2	1.513	0.382	0.252	0.281
Ghana	2005	GB2	1.344	(0.161)	2.863	(0.185)	2.392	(0.483)	2.276	(0.447)	-17,228.6	34,465.2	34,493.0	4.068	0.419	0.305	0.325
		D	2.253	(0.046)	2.838	(0.109)	1.083	(0.055)	—		-17,245.2*	34,496.4	34,517.2	4.202	0.438	0.335	0.394
		SM	2.298	(0.047)	3.035	(0.110)	—		1.022	(0.051)	-17,246.6*	34,499.3	34,520.1	4.140	0.430	0.326	0.371
Kenya	2005	GB2	1.378	(0.124)	2.317	(0.165)	2.275	(0.366)	1.384	(0.184)	-30,893.0	61,794.0	61,823.5	6.223	0.534	0.506	0.659
		D	1.706	(0.026)	2.467	(0.115)	1.619	(0.084)	—		-30,898.4*	61,802.8	61,824.9	6.518	0.556	0.550	0.793
		SM	2.263	(0.048)	2.806	(0.089)	—		0.697	(0.029)	-30,924.3*	61,854.7	61,876.8	6.770	0.574	0.593	0.940
Malawi	2013	GB2	1.927	(0.272)	1.536	(0.172)	2.650	(0.744)	1.326	(0.265)	-7,244.4	14,496.8	14,521.9	3.144	0.381	0.242	0.288
		D	2.337	(0.056)	1.663	(0.108)	1.888	(0.194)	—		-7,245.8	14,497.6	14,516.4	3.184	0.389	0.254	0.315
		SM	3.339	(0.136)	1.920	(0.073)	—		0.639	(0.046)	-7,257.4*	14,520.9	14,539.7	3.223	0.399	0.269	0.352
Malawi	2017	GB2	2.243	(0.192)	0.839	(0.054)	2.431	(0.415)	1.023	(0.120)	-15,472.3	30,952.6	30,982.1	1.852	0.390	0.254	0.322
		D	2.279	(0.036)	0.846	(0.037)	2.359	(0.164)	—		-15,472.3	30,950.7	30,972.8	1.854	0.391	0.255	0.325
		SM	3.675	(0.085)	1.041	(0.020)	—		0.545	(0.022)	-15,504.7*	31,015.3	31,037.4	1.903	0.409	0.282	0.393

Table 3: Continued

Country	Year	Model[a]	$\widehat{a_c}$		$\widehat{b_c}$		$\widehat{p_c}$		$\widehat{q_c}$		ln L	AIC	BIC	Mean	Gini	MLD	Theil
							Parameters[b]					**Comparison fit statistics[c]**			**Predictions[d]**			
Niger	2011	GB2	2.317	(0.458)	1.341	(0.231)	3.649	(1.616)	1.346	(0.359)	-5,521.0	11,049.9	11,074.6	2.688	0.301	0.148	0.171
		D	2.856	(0.080)	1.510	(0.106)	2.395	(0.323)		—	-5,522.2	11,050.3	11,068.8	2.706	0.306	0.154	0.182
		SM	4.560	(0.231)	1.810	(0.062)		—	0.559	(0.046)	-5,536.1*	11,078.2	11,096.7	2.729	0.314	0.165	0.205
Niger	2014	GB2	0.862	(0.418)	0.271	(0.680)	26.850	(54.380)	4.874	(3.442)	-5,486.0	10,980.1	11,004.5	2.632	0.344	0.194	0.211
		D	2.482	(0.082)	1.480	(0.133)	1.932	(0.275)		—	-5,498.7*	11,003.5	11,021.8	2.712	0.364	0.221	0.269
		SM	3.438	(0.183)	1.778	(0.097)		—	0.690	(0.069)	-5,515.1*	11,036.2	11,054.5	2.703	0.364	0.224	0.275
Nigeria	2004	GB2	1.043	(0.122)	1.959	(0.150)	3.751	(0.757)	3.977	(0.871)	-29,125.3	58,258.5	58,289.5	2.416	0.398	0.274	0.279
		D	2.363	(0.037)	1.797	(0.047)	1.038	(0.039)		—	-29,205.8*	58,417.6	58,440.8	2.511	0.420	0.308	0.354
		SM	2.300	(0.033)	1.989	(0.057)		—	1.127	(0.046)	-29,199.4*	58,404.8	58,428.0	2.454	0.408	0.293	0.318
Rwanda	2014	GB2	2.709	(0.182)	1.094	(0.045)	1.576	(0.195)	0.659	(0.055)	-23,914.4	47,836.9	47,867.2	2.719	0.471	0.377	0.570
		D	1.982	(0.026)	0.908	(0.040)	2.810	(0.178)		—	-23,928.8*	47,863.6	47,886.3	2.616	0.450	0.341	0.468
		SM	3.519	(0.059)	1.224	(0.017)		—	0.482	(0.014)	-23,925.2*	47,856.3	47,879.0	2.783	0.484	0.401	0.644
Tanzania	2009	GB2	1.430	(0.312)	0.793	(0.281)	5.610	(3.114)	1.858	(0.573)	-5,075.9	10,159.8	10,183.8	2.551	0.393	0.255	0.303
		D	2.186	(0.058)	1.152	(0.096)	2.314	(0.285)		—	-5,079.5*	10,165.0	10,183.1	2.623	0.410	0.281	0.364
		SM	3.397	(0.147)	1.456	(0.058)		—	0.582	(0.043)	-5,094.2*	10,194.3	10,212.4	2.667	0.422	0.303	0.419
Tanzania	2013	GB2	1.087	(0.283)	1.015	(0.395)	5.854	(3.548)	2.843	(1.130)	-8,476.8	16,961.6	16,987.3	2.859	0.409	0.281	0.313
		D	2.124	(0.051)	1.486	(0.103)	1.717	(0.166)		—	-8,489.0*	16,984.0	17,003.3	2.992	0.436	0.323	0.413
		SM	2.833	(0.108)	1.730	(0.075)		—	0.709	(0.050)	-8,506.1*	17,018.2	17,037.5	3.001	0.440	0.333	0.439
Uganda	2005	GB2	1.365	(0.169)	0.447	(0.119)	6.036	(2.151)	1.474	(0.257)	-11,375.7	22,759.3	22,786.8	2.295	0.480	0.389	0.522
		D	1.802	(0.031)	0.667	(0.045)	3.074	(0.263)		—	-11,379.5*	22,764.9	22,785.6	2.378	0.498	0.423	0.619
		SM	3.143	(0.073)	1.009	(0.024)		—	0.508	(0.021)	-11,418.3*	22,842.6	22,863.3	2.515	0.528	0.484	0.816

Source: World Bank.

Note: AIC = Akaike information criteria; BIC = Bayesian information criteria; MLD = mean log deviation.

a. GB2 = generalized beta II, D = Dagum, SM = Singh-Maddala.

b. Numbers in parentheses: estimated standard errors.

c. Asterisks placed next to the log-likelihood values of the Dagum and/or the Singh-Maddala distribution indicate that the improvement gained in adding a further parameter is of practical significance at the 5 percent level.

d. Analytic values obtained by substituting the estimated parameters into the relevant expressions—the formulas for the generalized beta II, Dagum, and Singh-Maddala distributions can be found in Kleiber and Kotz (2003, ch. 6) and Jenkins (2009).

Table 4: Maximum Likelihood Estimation of Generalized Beta Models for Income Distributions

Country	Year	Model[a]	$\widehat{a_c}$		$\widehat{b_c}$		$\widehat{p_c}$		$\widehat{q_c}$		ln L	AIC	BIC	Mean	Gini	MLD	Theil
							Parameters[b]					Comparison fit statistics[c]				Predictions[d]		
Burkina Faso	2014	GB2	2.284	(0.240)	0.323	(0.012)	0.680	(0.089)	0.647	(0.090)	-3,625.1	7,258.2	7,286.7	0.720	0.630	0.758	1.233
		D	1.692	(0.030)	0.338	(0.015)	0.995	(0.047)	—		-3,633.6*	7,273.2	7,294.6	0.653	0.591	0.662	0.887
		SM	1.719	(0.042)	0.323	(0.015)	—		0.952	(0.047)	-3,632.8*	7,271.5	7,292.9	0.670	0.602	0.684	0.956
Ethiopia	2014	GB2	2.704	(0.547)	0.519	(0.030)	0.401	(0.098)	0.588	(0.146)	-2,667.0	5,342.1	5,367.9	0.776	0.611	0.761	1.009
		D	1.865	(0.058)	0.583	(0.038)	0.621	(0.047)	—		-2,673.0*	5,352.0	5,371.4	0.718	0.580	0.685	0.761
		SM	1.327	(0.054)	0.629	(0.068)	—		1.620	(0.152)	-2,689.8*	5,385.6	5,405.0	0.709	0.574	0.662	0.689
Ethiopia	2016	GB2	2.140	(0.443)	0.521	(0.033)	0.575	(0.151)	0.663	(0.175)	-3,442.6	6,893.2	6,918.6	1.134	0.675	0.920	1.509
		D	1.608	(0.048)	0.557	(0.041)	0.822	(0.061)	—		-3,445.7*	6,897.4	6,916.5	1.016	0.637	0.812	1.084
		SM	1.416	(0.055)	0.544	(0.049)	—		1.175	(0.096)	-3,449.2*	6,904.4	6,923.4	1.010	0.634	0.801	1.034
Ghana	2005	GB2	0.716	(0.092)	0.961	(0.247)	2.160	(0.418)	4.085	(1.096)	-4,371.7	8,751.5	8,779.2	0.722	0.610	0.773	0.746
		D	1.507	(0.030)	0.443	(0.024)	0.786	(0.037)	—		-4,406.4*	8,818.8	8,839.6	0.882	0.682	0.968	1.362
		SM	1.208	(0.027)	0.571	(0.048)	—		1.547	(0.098)	-4,387.2*	8,780.4	8,801.2	0.766	0.634	0.837	0.926
Kenya	2005	GB2	1.060	(0.096)	2.517	(0.214)	1.159	(0.148)	1.537	(0.232)	-26,723.6	53,455.3	53,484.7	4.804	0.695	1.036	1.290
		D	1.377	(0.025)	2.190	(0.105)	0.821	(0.033)	—		-26,732.0*	53,470.0	53,492.1	5.612	0.739	1.189	1.913
		SM	1.174	(0.022)	2.428	(0.167)	—		1.308	(0.067)	-26,724.9	53,455.8	53,477.9	4.973	0.706	1.072	1.428
Malawi	2013	GB2	1.377	(0.191)	0.711	(0.063)	1.156	(0.233)	1.359	(0.289)	-3,912.2	7,832.3	7,857.4	1.164	0.588	0.668	0.794
		D	1.669	(0.046)	0.670	(0.046)	0.893	(0.061)	—		-3,913.7	7,833.4	7,852.2	1.223	0.608	0.716	0.946
		SM	1.522	(0.050)	0.702	(0.056)	—		1.169	(0.088)	-3,912.6	7,831.2	7,849.9	1.182	0.594	0.685	0.846
Malawi	2017	GB2	2.683	(0.256)	0.439	(0.013)	0.474	(0.055)	0.552	(0.066)	-6,682.5	13,373.1	13,402.6	0.839	0.632	0.785	1.227
		D	1.756	(0.032)	0.477	(0.016)	0.797	(0.031)	—		-6,705.0*	13,416.0	13,438.1	0.748	0.588	0.674	0.835
		SM	1.531	(0.030)	0.451	(0.020)	—		1.165	(0.054)	-6,722.2*	13,451.4	13,473.6	0.758	0.592	0.678	0.837

Table 4: Continued

Country	Year	Model[a]	$\widehat{a_c}$		$\widehat{b_c}$		$\widehat{p_c}$		$\widehat{q_c}$		ln L	AIC	BIC	Mean	Gini	MLD	Theil
							Parameters[b]					Comparison fit statistics[c]			Predictions[d]			
Niger	2011	GB2	1.324	(0.219)	0.730	(0.123)	0.760	(0.160)	1.547	(0.425)	-1,896.8	3,801.6	3,826.3	0.706	0.617	0.826	0.816
		D	1.701	(0.059)	0.593	(0.046)	0.562	(0.041)		—	-1,899.1*	3,804.2	3,822.7	0.753	0.641	0.889	1.010
		SM	1.077	(0.042)	0.903	(0.154)		—	2.254	(0.268)	-1,898.3	3,802.6	3,821.2	0.690	0.608	0.801	0.747
Niger	2014	GB2	1.417	(0.260)	0.601	(0.105)	0.745	(0.174)	1.373	(0.412)	-1,513.5	10,980.1	11,004.5	0.658	0.614	0.802	0.832
		D	1.704	(0.061)	0.523	(0.046)	0.597	(0.051)		—	-1,514.7	11,003.5	11,021.8	0.692	0.633	0.852	0.986
		SM	1.134	(0.054)	0.726	(0.135)		—	2.029	(0.276)	-1,515.1	11,036.2	11,054.5	0.602	0.602	0.771	0.744
Nigeria	2004	GB2	1.537	(0.125)	3.480	(0.253)	0.588	(0.060)	1.271	(0.177)	-35,254.4	70,516.9	70,547.8	3.336	0.620	0.857	0.841
		D	1.785	(0.039)	3.170	(0.097)	0.490	(0.017)		—	-35,258.6*	70,523.2	70,546.4	3.436	0.630	0.886	0.926
		SM	1.040	(0.015)	4.672	(0.448)		—	2.411	(0.176)	-35,282.6*	70,571.1	70,594.3	3.257	0.613	0.828	0.756
Rwanda	2014	GB2	1.170	(0.085)	0.725	(0.037)	1.311	(0.140)	1.408	(0.162)	-17,327.9	34,663.8	34,694.1	1.586	0.658	0.871	1.137
		D	1.448	(0.019)	0.677	(0.025)	0.984	(0.034)		—	-17,334.1*	34,674.1	34,696.8	1.757	0.692	0.972	1.520
		SM	1.410	(0.023)	0.709	(0.030)		—	1.060	(0.039)	-17,332.7*	34,671.3	34,694.0	1.689	0.680	0.936	1.382
Tanzania	2009	GB2	0.788	(0.153)	1.041	(0.340)	1.797	(0.522)	3.136	(1.209)	-2,619.5	5,246.9	5,271.0	0.995	0.629	0.834	0.827
		D	1.461	(0.045)	0.577	(0.053)	0.788	(0.061)		—	-2,630.0*	5,266.1	5,284.1	1.231	0.701	1.041	1.521
		SM	1.180	(0.045)	0.729	(0.103)		—	1.509	(0.154)	-2,623.7*	5,253.4	5,271.4	1.060	0.653	0.901	1.028
Tanzania	2013	GB2	0.861	(0.139)	1.222	(0.266)	1.819	(0.433)	2.691	(0.824)	-5,822.1	11,652.2	11,677.9	1.484	0.614	0.768	0.796
		D	1.485	(0.034)	0.799	(0.058)	0.881	(0.056)		—	-5,833.7*	11,673.5	11,692.8	1.791	0.682	0.952	1.404
		SM	1.305	(0.045)	0.934	(0.098)		—	1.296	(0.107)	-5,828.7*	11,663.4	11,682.7	1.608	1.608	0.852	1.053
Uganda	2005	GB2	2.197	(0.225)	0.566	(0.020)	0.697	(0.092)	0.604	(0.081)	-7,951.0	15,910.0	15,937.5	1.719	0.702	0.975	1.955
		D	1.538	(0.028)	0.580	(0.026)	1.110	(0.051)		—	-7,960.1*	15,926.3	15,946.9	1.436	0.643	0.798	1.191
		SM	1.692	(0.036)	0.552	(0.023)		—	0.851	(0.038)	-7,955.9*	15,917.8	15,938.5	1.541	0.667	0.863	1.441

Source: World Bank.

a. GB2 = generalized beta II; D = Dagum; AIC = Akaike information criteria; BIC = Bayesian information criteria; MLD = mean log deviation; SM = Singh-Maddala.

b. Numbers in parentheses: estimated standard errors.

c. Asterisks placed next to the log-likelihood values of the Dagum and/or the Singh-Maddala distribution indicate that the improvement gained in adding a further parameter is of practical significance at the 5 percent level.

d. Analytic values obtained by substituting the estimated parameters into the relevant expressions—the formulas for the generalized beta II, Dagum, and Singh-Maddala distributions can be found in Kleiber and Kotz (2003, ch. 6) and Jenkins (2009).

Table 5: Inequality Estimates from Partially Synthetic Datasets, by Country/Year, Definition of the Middle Class and Index

Country	Year	Model[a] Consumption	Income	$5.5/day Middle class size (%)	Gini[b]		MLD[b]		Theil[b]		$10/day Middle class size (%)	Gini[b]		MLD[b]		Theil[b]	
Burkina Faso	2014	D	GB2	8.060	0.469	(0.031)	0.384	(0.068)	0.746	(0.255)	1.960	0.418	(0.028)	0.301	(0.058)	0.558	(0.222)
Ethiopia	2014	SM	GB2	2.230	0.431	(0.026)	0.326	(0.051)	0.501	(0.165)	0.690	0.420	(0.029)	0.311	(0.071)	0.467	(0.187)
Ethiopia	2016	GB2	GB2	1.300	0.419	(0.025)	0.308	(0.052)	0.481	(0.183)	0.150	0.381	(0.015)	0.248	(0.025)	0.302	(0.099)
Ghana	2005	GB2	GB2	20.260	0.455	(0.008)	0.359	(0.013)	0.414	(0.029)	5.880	0.438	(0.007)	0.334	(0.012)	0.382	(0.027)
Kenya	2005	GB2	SM	31.370	0.616	(0.024)	0.691	(0.074)	1.075	(0.256)	14.190	0.608	(0.023)	0.671	(0.067)	1.047	(0.239)
Malawi	2013	D	SM	10.050	0.445	(0.020)	0.338	(0.036)	0.510	(0.117)	2.620	0.411	(0.017)	0.289	(0.029)	0.420	(0.100)
Malawi	2017	GB2	GB2	66.500	0.553	(0.027)	0.532	(0.076)	0.947	(0.264)	28.690	0.540	(0.029)	0.510	(0.078)	0.925	(0.274)
Niger	2011	D	SM	5.850	0.332	(0.011)	0.185	(0.014)	0.246	(0.035)	1.030	0.311	(0.007)	0.160	(0.008)	0.201	(0.022)
Niger	2014	GB2	D	6.670	0.417	(0.024)	0.299	(0.045)	0.473	(0.144)	1.160	0.372	(0.014)	0.234	(0.023)	0.328	(0.086)
Nigeria	2004	GB2	GB2	6.600	0.451	(0.011)	0.356	(0.021)	0.466	(0.075)	1.090	0.415	(0.009)	0.300	(0.016)	0.353	(0.059)
Rwanda	2014	GB2	GB2	8.120	0.513	(0.019)	0.453	(0.041)	0.754	(0.170)	2.750	0.493	(0.018)	0.417	(0.038)	0.677	(0.157)
Tanzania	2009	GB2	GB2	7.300	0.413	(0.012)	0.285	(0.018)	0.359	(0.044)	1.670	0.401	(0.009)	0.268	(0.014)	0.329	(0.035)
Tanzania	2013	GB2	GB2	10.580	0.451	(0.012)	0.344	(0.020)	0.436	(0.053)	2.370	0.427	(0.010)	0.309	(0.016)	0.375	(0.045)
Uganda	2005	GB2	GB2	6.380	0.582	(0.035)	0.607	(0.109)	1.174	(0.358)	2.000	0.542	(0.034)	0.518	(0.094)	0.966	(0.332)

Source: World Bank.

Note: the numbers in parentheses denote the estimated standard errors derived using the methods discussed in section 4.1. MLD = mean log deviation.

a. GB2 = generalized beta II; D = Dagum; SM = Singh-Maddala.

b. Average of the point estimates derived from each of the R = 1,000 partially synthetic data set.

References

Acemoglu, D., and J. A. Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. New York: Crown Business.

Adhvaryu, A., J. Fenske, G. Khanna, A. Nyshadham. 2021. "Resources, Conflict, and Economic Development in Africa." *Journal of Development Economics* 149: 102598.

African Development Bank. 2011. "The Middle of the Pyramid: Dynamics of the Middle Class in Africa." Market brief, African Development Bank, Abidjan. https://www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/The%20Middle%20of%20the%20Pyramid_The%20Middle%20of%20the%20Pyramid.pdf.

Aguiar, M., and M. Bils. 2018. "Has Consumption Inequality Mirrored Income Inequality?" *American Economic Review* 105: 2725–56.

Akaike, H. 1973. "Information Theory and an Extension of the Likelihood Ratio Principle." In *Proceedings of the Second International Symposium of Information Theory*, edited by B. N. Petrov, and F. Csaki, 257–81. Budapest: Akademiai Kiado.

Alderson, A., and F. Nielsen. 2002. "Globalization and the Great U-Turn: Income Inequality Trends in 16 OECD Countries." *American Journal of Sociology* 107: 1244–99.

Alstadsæter, A., N. Johannesen, and G. Zucman. 2018. "Who Owns the Wealth in Tax Havens? Macroevidence and Implications for Global Inequality." *Journal of Public Economics* 162: 89–100.

Alvaredo, F. 2010. "The Rich in Argentina over the Twentieth Century: 1932–2004." In *Top Incomes: A Global Perspective*, edited by A. B. Atkinson and T. Piketty, 253–98. Oxford, UK: Oxford University Press.

An, D., and R. J. A. Little. 2007. "Multiple Imputation: An Alternative to Top Coding for Statistical Disclosure Control." *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 170: 923–40.

Attanasio, O. P., and L. Pistaferr1. 2016. "Consumption Inequality." *Journal of Economic Perspectives* 30: 2–28.

Atkinson A. B. 2014 "The Colonial Legacy: Income Inequality in Former British African Colonies." WIDER Working Paper 45, United Nations University World Institute for Development Economic Research, Helsinki. https://www.wider.unu.edu/publication/colonial-legacy-0.

Atkinson, A. B., T. Piketty, and E. Saez. 2011. "Top Incomes in the Long Run of History." *Journal of Economic Literature* 49: 3–71.

Azizi, S. 2021. "The Impacts of Workers' Remittances on Poverty and Inequality in Developing Countries." *Empirical Economics* 60: 969–91.

Beegle, K., L. Christiaensen, A. L. Dabalen, and I. Gaddis. 2016. *Poverty in a Rising Africa*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/22575.

Bhorat, H., K. Naidoo, and K. Pillay. 2016. "Growth, Poverty and Inequality Interactions in Africa: An Overview of Key Issues." UNDP-RBA Working Paper 1, United Nations Development Programme (UNDP) Regional Bureau for Africa (RBA), New York. https://ageconsearch.umn.edu/record/267778/files/RBA_WPS_1_Growth%2C%20Poverty%20an

d%20Inequality%20Interactions%20in%20Africa%20%20An%20Overview%20of.%20Key%20Issues.pdf.

Bhorat, H., R. Kanbur, and F. K. Steenkamp. 2017. *Sub-Saharan Africa's Manufacturing Sector: Building Complexity*. CEPR Discussion Paper 12073, Center for Economic and Policy Research, Washington, DC, 2017. https://cepr.org/active/publications/discussion_papers/dp.php?dpno=12073.

Bourguignon, F. 2004. "The Poverty-Growth-Inequality Triangle." Working Paper 125, Indian Council for Research on International Economic Relations, New Delhi. http://www.icrier.org/pdf/wp125.pdf.

Brzozowski, M., M. Gervais, P. Klein, M. Suzuki. 2010. "Consumption, Income, and Wealth Inequality in Canada." *Review of Economic Dynamics* 13: 52–75.

Carletto, G., K. Covarrubias, B. Davis, M. Krausova, and P. Winters. 2007. *Rural Income Generating Activities Study: Methodological Note on the Construction of Income Aggregates*. Prepared for the Rural Income Generating Activities (RIGA) project of the Agricultural Development Economics Division, Food and Agriculture Organization. http://www.fao.org/fileadmin/user_upload/riga/pdf/ai197e00.pdf.

Chancel, L., D. Cogneau, A. Gethin, and A. Myczkowski. 2019. "How Large Are African Inequalities ? Towards Distributional National Accounts in Africa, 1990–2017." WID.world Working Paper 13, World Inequality Lab, Paris School of Economics and the University of California Berkeley. https://wid.world/document/cgm2019-full-paper/.

Christiaensen, L., P. Chuhan-Pole, and A. Sanoh. 2013. "Africa's Growth, Poverty, and Inequality Nexus—Fostering Shared Prosperity." Internal document draft, World Bank, Washington, DC. https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=CSAE2014&paper_id=381.

Clementi F., M Fabiani, V Molini (2020) "How polarized is sub-Saharan Africa? A look at the regional distribution of consumption expenditure in the 2000s" *Oxford Economic Papers* 73 (2), 796–819

Clementi, F., A. L. Dabalen, V. Molini, and F. Schettino. 2020. "We Forgot the Middle Class! Inequality Underestimation in a Changing Sub-Saharan Africa." *Journal of Economic Inequality* 18: 45–70.

Cornia, G. A. 2017. "Inequality Levels, Trends, and Determinants in Sub-Saharan Africa: An Overview of Main Changes since the Early 1990s." In *Income Inequality Trends in Sub-Saharan Africa: Divergence, Determinants and Consequences*, edited by A. Odusola, G. A. Cornia, H. Bhorat, and P. Conceição, 23–51. New York: United Nations Development Programme.

Corral Rodas, P. A., V. Molini, and G. Oseni. 2019. "No Condition is Permanent: Middle Class in Nigeria in the Last Decade." *Journal of Development Studies* 55: 294–310.

Dagum, C. 1977. "A New Model of Personal Income Distribution: Specification and Estimation." *Economie Appliquée* 30: 413–36.

De Magalhães, L.,and R. Santaeulàlia-Llopis. 2018. "The Consumption, Income, and Wealth of the Poorest: An Empirical Analysis of Economic Inequality in Rural and Urban Sub-Saharan Africa for Macroeconomists." *Journal of Development Economics* 134: 350–71.

Deaton, A., and S. Zaidi. 2002. "Guidelines for Constructing Consumption Aggregates for Welfare Analysis." LSMS Working Paper 135, World Bank, Washington DC. https://openknowledge.worldbank.org/handle/10986/14101.

Devarajan, S. 2013. "Africa's Statistical Tragedy." *Review of Income and Wealth* 59: 9–15.

Deville, J.-C., C.-E. Särndal, and O. Sautory. 1993. "Generalized Raking Procedures in Survey Sampling." *Journal of the American Statistical Association* 88: 1013–1020.

Devarajan, S., and M. Giugal. 2013 *The Case for Direct Transfers of Resource Revenues in Africa*. CGD Working Paper 333, Washington, DC, 2013. https://www.cgdev.org/publication/case-direct-transfers-resource-revenues-africa-working-paper-333.

Diao, X., E. Magalhaes, and M., McMillan. 2018. "Understanding the Role of Rural Non-Farm Enterprises in Africa's Economic Transformation: Evidence from Tanzania." *Journal of Development Studies* 54: 833–55.

Ding, H., and H. He. 2018. "A Tale of Transition: An Empirical Analysis of Economic Inequality in Urban China, 1986‒2009." *Review of Economic Dynamics* 29: 106–37.

Dynan, K. E., J. S. Skinner, and S. P. Zeldes. 2004. "Do the Rich Save More?" *Journal of Political Economy* 112: 397–444.

Fisher, J. D., D. S. Johnson, and T. M. Smeeding. 2015. "Inequality of Income and Consumption in the U.S.: Measuring the Trends in Inequality from 1984 to 2011 for the Same Individuals." *Review of Income and Wealth* 61: 630–50.

Fisher, J. D., D. S. Johnson, J. P. Latner, T. M. Smeeding, and J. P. Thompson. 2016. "Inequality and Mobility Using Income, Consumption, and Wealth for the Same Individuals." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2: 44–58.

Fisher, J. D., D. S. Johnson, T. M. Smeeding, and J. P. Thompson. 2021. "Inequality in 3-D: Income, Consumption, and Wealth." *Review of Income and Wealth*, https://doi.org/10.1111/roiw.12509.

Förster, M. F., and I. G. Tóth. 2015. "Cross-Country Evidence of the Multiple Causes of Inequality Changes in the OECD Area." In *Handbook of Income Distribution,* vol. 2, edited by A. B. Atkinson and F. Bourguignon , 1729–1843. Amsterdam: North-Holland.

Fosu, A. K. 2009. "Inequality and the Impact of Growth on Poverty: Comparative Evidence for Sub-Saharan Africa." *Journal of Development Studies* 45: 726–45.

Fosu, A. K. 2017a. "Growth, Inequality, and Poverty Reduction in Developing Countries: Recent Global Evidence." *Research in Economics* 71: 306–36.

Fosu, A. K. 2017b. "Growth, Inequality, and Poverty Reduction: Africa in a Global Setting." In *Poverty Reduction in the Course of African Development*, edited by M. Nissanke and M. Ndulo, 57–76. Oxford, UK: Oxford University Press.

Fosu, A. K. 2018. "The Recent Growth Resurgence in Africa and Poverty Reduction: The Context and Evidence." *Journal of African Economies* 27: 92–107.

Frazis, H., Stewart, J. (2011) How does household production affect measured income inequality? *Journal of Population Economics* 24, 3–22 (2011).

Friedman, M. 1957. *A Theory of the Consumption Function*. Princeton, NJ: Princeton University Press.

Gandelman, N. 2017. "Do the Rich Save More in Latin America?" *Journal of Economic Inequality* 15: 75–92.

Gollin, D., D. Lagakos, and M. E. Waugh. 2014. "The Agricultural Productivity Gap." *Quarterly Journal of Economics* 129: 939–93.

Harttgen, K., S. Klasen, and S. Vollmer. 2013. "An African Growth Miracle? Or: What do Asset Indices Tell Us about Trends in Economic Performance?" *Review of Income and Wealth* 59: 37–61.

Heathcote, J., F. Perri, and G. L. Violante. 2010. "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967–2006." *Review of Economic Dynamics* 13: 15–51.

ILO (International Labour Organization). 2003.

Hlasny, V., and P. Verme. 2021. "The Impact of Top Incomes Biases on the Measurement of Inequality in the United States." *Oxford Bulletin of Economics and Statistics*, https://doi.org/10.1111/obes.12472.

Jappelli, T., and L. Pistaferri. 2010. "Does Consumption Inequality Track Income Inequality in Italy?" *Review of Economic Dynamics* 13: 133–53.

Jappelli, T., and L. Pistaferri. 2014. "Fiscal Policy and MPC Heterogeneity." *American Economic Journal: Macroeconomics* 6: 107–36.

Jenkins, S. P. 1999. "Fitting Singh-Maddala and Dagum Distributions by Maximum Likelihood." *Stata Technical Bulletin* 48: 19–25.

Jenkins, S. P. 2007. *GB2FIT: Stata Module to Fit Generalized Beta of the Second Kind Distribution by Maximum Likelihood*. Statistical Software Components Archive S456823. Federal Reserve Bank of St. Louis. http://ideas.repec.org/c/boc/bocode/s456823.

Jenkins, S. P. 2009. "Distributionally Sensitive Inequality Indices and the GB2 Income Distribution." *Review of Income and Wealth* 55: 392–98.

Jenkins, S. P. 2014. *GB2LFIT: Stata Module to Fit Generalized Beta of the Second Kind Distribution by Maximum Likelihood (Log Parameter Metric*. Statistical Software Components Archive S457897, Federal Reserve Bank of St. Louis. https://ideas.repec.org/c/boc/bocode/s457897.html.

Jenkins, S. P., R. V. Burkhauser, S. Feng, and J. Larrimore. 2011. "Measuring Inequality Using Censored Data: A Multiple-Imputation Approach to Estimation and Inference." *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 174: 63–81.

Jerven, M. 2015. *Africa: Why Economists Get It Wrong*. London: Zed Books.

Kanbur, R. 2017. "Structural Transformation and Income Distribution: Kuznets and Beyond." IZA Discussion Paper 10636, Institute of Labor Economics IZA, Bonn. http://ftp.iza.org/dp10636.pdf.

Kharas, H. 2010. "The Emerging Middle Class in Developing Countries." OECD Development Centre Working Paper 285, Organisation for Economic Co-operation and Development, Paris. https://doi.org/10.1787/5kmmp8lncrns-en.

Kleiber, C. 1996. "Dagum vs. Singh-Maddala Income Distributions." *Economics Letters* 53: 265–68.

Kleiber, C. 2008. "A Guide to the Dagum Distributions." in *Modeling Income Distributions and Lorenz Curves*, edited by D. Chotikapanich, 97–117. New York: Springer.

Kleiber, C., and S. Kotz. 2003. *Statistical Size Distributions in Economics and Actuarial Sciences*. New York, NYC: John Wiley & Sons.

Knight Frank Research. 2015. *The Wealth Report 2015: Global Perspectives on Prime Property and Wealth*. London: Citi Private Bank. https://www.knightfrank.com/research/the-wealth-report-2015-2716.aspx.

Krueger, D., and F. Perri. 2006. "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory." *Review of Economic Studies* 73: 163–93.

Kuznets, S. 1955. "Economic Growth and Income Inequality." *American Economic Review* 45: 1–28.

Leigh, A., and P. Van der Eng. 2009. "Inequality in Indonesia: What Can We Learn from Top Incomes?" *Journal of Public Economics* 93: 209–12.

Lewis, W. A. 1955. *The Theory of Economic Growth*. London: Allen & Unwin.

McCarthy, J. 1995. "Imperfect Insurance and Differing Propensities to Consume Across Households." *Journal of Monetary Economics* 36: 301–27.

McDonald, J. B. 1984. "Some Generalized Functions for the Size Distribution of Income." *Econometrica* 52: 647–65.

McDonald, J. B., and M. Ransom. 2008. "The Generalized Beta Distribution as a Model for the Distribution of Income: Estimation of Related Measures of Inequality." In *Modeling Income Distributions and Lorenz Curves*, edited by D. Chotikapanich, 147–66. New York: Springer.

McDonald, J. B., and Y. J. Xu. 1995a. "A Generalization of the Beta Distribution with Applications." *Journal of Econometrics* 66: 133–52.

McDonald, J. B., and Y. J. Xu. 1995b. "Errata." *Journal of Econometrics* 69: 427–28.

McMillan, M., D. Rodrik, and Í. Verduzco-Gallo. 2014. "Globalization, Structural Change, and Productivity Growth, with an Update on Africa." *World Development* 63: 11–32.

Meyer, B. D., and J. X. Sullivan. 2004. "The Effects of Welfare and Tax Reform: The Material Well-Being of Single Mothers in the 1980s and 1990s." *Journal of Public Economics* 88: 1387–1420.

Meyer, B. D., and J. X. Sullivan. 2011. "Viewpoint: Further Results on Measuring the Well-Being of the Poor Using Income and Consumption." *Canadian Journal of Economics* 44: 52–87.

Milanovic, B. 2010. *The Haves and the Have-Nots: A Brief and Idiosyncratic History of Global Inequality*. New York: Basic Books.

Modigliani, F., snd R. H. Brumberg. 1954. "Utility Analysis and the Consumption Function: An Interpretation of Cross-Section Data." In *Post Keynesian Economics*, edited by K. K. Kurihara, 388–436. New Brunswick, NJ: Rutgers University Press.

Modigliani, F., and R. H. Brumberg. 1980. "Utility Analysis and Aggregate Consumption Functions: An Attempt at Integration." In *The Collected Papers of Franco Modigliani,* vol. 2, *The Life Cycle Hypothesis of Saving*, edited by A. Abel, 128–97. Cambridge, MA: MIT Press.

Molini, V., and P. Paci. 2015. *Poverty Reduction in Ghana: Progress and Challenges*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/22732.

Ncube, M., and C. L. Lufumpa, eds. 2014. *The Emerging Middle Class in Africa*. London: Routledge.

Nielsen, F., and A. Alderson. 1995. "Income Inequality, Development, and Dualism: Results from an Unbalanced Cross-National Panel." *American Sociological Review* 60: 674–701.

Odusola, A. F., G. A. Cornia, H. Bhorat, and P. Conceição, eds. 2017. *Income Inequality Trends in Sub-Saharan Africa: Divergence, Determinants and Consequences*. New York: United Nations Development Programme.

Piketty, T. 2014. *Capital in the Twenty-first Century*. Cambridge, MA: Belknap Press of Harvard University Press.

Piketty, T. 2020. *Capital and Ideology*. Cambridge, MA: Harvard University Press.

Pinkovskiy, M., and M. Sala-i-Martin. 2014. "Africa Is on Time." *Journal of Economic Growth* 19: 311–38.

Quiñones, E. J., A. P. de la O-Campos, C. Rodríguez-Alas, T. Hertz, and P. Winters. 2009. *Methodology for Creating the RIGA-L Database*. Prepared for the Rural Income Generating Activities (RIGA) project of the Agricultural Development Economics Division, Food and Agriculture Organization. http://www.fao.org/fileadmin/templates/riga/docs/Country_survey_information/RIGA-L_Methodology.pdf.

Ravallion, M. 2005. "A Poverty-Inequality Trade off?" *Journal of Economic Inequality* 3: 169–81.

Ravallion, M. 2010. "The Developing World's Bulging (but Vulnerable) Middle Class." *World Development* 38: 445–54.

Reiter, J. P. 2003. "Inference for Partially Synthetic, Public-Use Microdata Sets." *Survey Methodology* 29: 181–88.

Robinson, J. A., R. Torvik, and T. Verdier. 2006. "Political Foundations of the Resource Curse." *Journal of Development Economics* 79: 447–68.

Rodrik, D. 2016a. "An African Growth Miracle?" *Journal of African Economies* 27: 1–18.

Rodrik, D. 2016b. "Premature Deindustrialization." *Journal of Economic Growth* 21: 1–33.

Sala-i-Martin, X., and A. Subramanian. 2003. "Addressing the Natural Resource Curse: An Illustration from Nigeria." IMF Working Paper 139, International Monetary Fund, Washington, DC. https://www.imf.org/en/Publications/WP/Issues/2016/12/30/Addressing-the-Natural-Resource-Curse-An-Illustration-From-Nigeria-16582.

Sanhueza, C., and R. Mayer. 2011. "Top Incomes in Chile Using 50 Years of Household Surveys: 1957–2007." *Estudios de Economía* 38: 169–93.

Schwarz, G. E. 1978. "Estimating the Dimension of a Model." *The Annals of Statistics* 6: 461–64.

Schettino, F., and H. A. Khan. 2020. "Income Polarization in the USA: What Happened to the Middle Class in the Last Few Decades?" *Structural Change and Economic Dynamics* 53: 149–61.

Schluter, C., and M. Trede. 2002. "Tails of Lorenz Curves." *Journal of Econometrics* 109: 151–66.

Schotte, S., R. Zizzamia, and M. Leibbrandt. 2018. "A Poverty Dynamics Approach to Social Stratification: The South African Case." *World Development* 110: 88–103.

Shimeles, A., and M. Ncube. 2015. "The Making of the Middle-Class in Africa: Evidence from DHS Data." *Journal of Development Studies* 51: 178–93.

Singh, S. K., and G. S. Maddala. 1976. "A Function for Size Distribution of Incomes." *Econometrica* 44: 963–70.

Spinesi, L. 2009. "Rent-Seeking Bureaucracies, Inequality, and Growth." *Journal of Development Economics* 90: 244–57.

Spilimbergo, A., J. L. Londoño, and M. Székely 1999. "Income Distribution, Factor Endowments, and Trade Openness." *Journal of Development Economics* 59: 77–101.

Tan, L. 2021. "Imputing Top-Coded Income Data in Longitudinal Surveys." *Oxford Bulletin of Economics and Statistics* 83: 66–87.

Tarozzi, A. 2007. "Calculating Comparable Statistics from Incomparable Surveys, with An Application to Poverty in India." *Journal of Business & Economic Statistics* 25: 314–36.

Thorbecke, E., and Y. Ouyang. 2018. "Is the Structure of Growth Different in Sub-Saharan Africa?" *Journal of African Economies* 27: 1–26.

World Bank. 2019. *Structural Transformation in Sub-Saharan Africa*. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/33327.

Zaidi, M., de Vos, K. 2001. Trends in consumption-based poverty and inequality in the European Union during the 1980s. *Journal of Population Economics* 14, 367–390 (2001).