the review of income and wealth

Review of Income and Wealth
Series 68, Number S2, December 2022
DOI: 10.1111/roiw.12577

DISPARITIES IN ASSESSMENTS OF LIVING STANDARDS USING NATIONAL ACCOUNTS AND HOUSEHOLD SURVEYS

Espen Beer Prydz*

The World Bank, Development Economics Data Group (DECDG)
The Norwegian Institute of Public Health, Center for Fertility and Health

Dean Jolliffe

The World Bank, Development Economics Data Group (DECDG)
IZA, Global Labor Organization (GLO), Johns Hopkins—SAIS

AND

Umar Serajuddin

The World Bank, Development Economics Data Group (DECDG)

Estimates of average per capita consumption and income from national accounts differ substantially from corresponding measures of consumption and income from household surveys. Using a new compilation of more than 2,000 household surveys matched to national accounts data, we find that the gaps between the data sources are larger and more robust than previously established. Means of household consumption estimated from surveys are, on average, 20 percent lower than corresponding means from national accounts. The gap with GDP per capita is nearly 50 percent. The gaps have increased in recent decades and are largest in middle-income countries, where annualized growth rates for consumption surveys are systematically lower than national accounts growth rates. We show that the gaps in measures across these two sources have implications for assessments of economic growth, poverty, and inequality. We find that typical survey measures of consumption and income may exaggerate poverty reduction and underestimate inequality.

JEL Codes: I3, I32, E31

Keywords: national accounts systems, household income and expenditure surveys, poverty, inequality

Note: The authors appreciate feedback from Daniel Mahler, Christoph Lakner, Marshall Reinsdorf and Xavier Mancero, as well as conference participants at the Special IARIW-World Bank Conference, Washington, DC. The authors gratefully acknowledge financial support from: the UK Government through the Knowledge for Change and the Data and Evidence for Tackling Extreme Poverty (DEEP) Programs, the Republic of Korea, acting through the Korea Development Institute School of Public Policy and Management (KDIS). The analysis in this paper would not be possible without data from national accounts and household surveys carried out by national governments around the world. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of any of the funders, author affiliations, or the World Bank Group and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent.

*Correspondence to: Espen Beer Prydz, The World Bank, Development Economics Data Group (DECDG) & The Norwegian Institute of Public Health, Center for Fertility and Health, 1818 H Street, N. W. Washington, DC 20433 USA (espen.prydz@gmail.com).

1. Introduction

Measures of per capita income and consumption are among the most frequently cited indicators of economic development. They are widely used in assessments of living standards, economic growth, poverty, and inequality, both within and across countries. Despite their prominence, the two most common data sources for such measures—national accounts systems (NAS) and household surveys (HHS)—often have large gaps between them and offer differing portrayals of living standards and economic growth. For example, for Pakistan in 2015, national accounts data suggest that average household consumption expenditure per capita was $9.3 per day at 2011 PPP, while the household survey indicates it was just a bit more than half of that, $4.9 per day at 2011 PPP.[1] In Botswana, the two recent household surveys suggest that per capita consumption contracted at an annualized rate of −3.3 percent between 2009 and 2015, while the most closely aligned measure from the national accounts system, per capita household final consumption expenditures (HFCE), indicated a robust expansion of household consumption at an annualized rate of 3.7 percent over the same period, as did gross domestic product (GDP).[2]

That national accounts data and survey data can lead to such diverging measures of the levels and changes in living standards is a recurring phenomenon across a wide range of countries and statistical systems. A frequently cited case is India, where large discrepancies in measures of household consumption expenditures across the national accounts and the National Sample Survey (NSS) have fueled a vigorous debate about the evolution of poverty and its relationship to economic growth (see for example Sundaram and Tendulkar, 2003; Deaton and Kozel, 2005; Subramanian and Jayaraj, 2015). The issue of diverging estimates from national accounts and household surveys is not limited to less wealthy countries. In the United States, per capita income from the two large national surveys, the Current Population Survey and the Consumer Expenditure Survey, are known to diverge from the national accounts (see for example the recent assessment by Schündeln, 2018). Nolan *et al*. (2019) recently reviewed gaps between survey incomes and national accounts incomes in 27 OECD countries, finding discrepancies in annual growth rates of 0.32 percentage points in the United States and 0.55 percentage points in Germany.

A number of studies have systematically reviewed the discrepancies between survey and national accounts data. The most complete reviews were conducted by Ravallion (2003), Karshenas (2001, 2003) and Deaton (2005), who all assess the discrepancies globally with a sample of household surveys and national accounts data from 1980s and 1990s. They document significant HHS–NAS gaps, but their

[1]Survey data for this example comes from the World Bank, Global Database of Shared Prosperity (GDSP) circa 2013–2018 (see www.worldbank.org/en/topic/poverty/brief/global-database-of-shared-prosperity) and national accounts data comes from the World Development Indicators, from latest data using the 2011 PPPs.

[2]Estimates are from the 2009 to 2015 spell for Botswana in the Global Database of Shared Prosperity (GDSP) and growth is estimated in constant terms. The survey years are decimal years (2009.25 and 2015.85) and national accounts growth is estimated for the same period using weighted annual data.

estimates varied substantially from each other and they were limited by relatively small samples of matched NAS and HHS data. Since these global assessments, the availability of household survey data in poorer countries have expanded considerably and many countries have revised both their survey and national accounts data and methods. We now also have better metadata on the types of household surveys and their comparability, which can help us better understand discrepancies.

The main contributions of this paper are threefold. We compile a data set of 2,095 household survey means from 166 countries matched to corresponding national accounts aggregates. Using this much larger compilation of matched NAS and HHS data, with broader geographic coverage and with temporal coverage stretching from 1965 to 2019, we revisit the HHS–NAS gaps and update the findings of Ravallion (2003), Karshenas (2003) and Deaton (2005). Second, we overcome the heterogeneity often seen in household surveys by extending the analysis to a subsample of the data containing only comparable spells, which is particularly important for understanding the discrepancies in estimates of growth. Third, we illustrate how the discrepancies in levels can lead to divergent representations of how living standards, poverty, and inequality differ across countries and over time.

Our findings suggest that disparities in income and consumption measures between surveys and national accounts are much larger than found by both Ravallion (2003) and Deaton (2005). On average, across all countries, we find that per capita consumption means are about 22 percent lower in surveys compared with national accounts (i.e., HFCE), and per capita income means from surveys are about 52 percent lower than per capita GDP. This indicates considerably larger discrepancies than what was found in Deaton's and Ravallion's assessments. Ravallion found no statistically significant gap in the means of household consumption across surveys and national accounts, while household income means from surveys were 33 percent lower than GDP in national accounts. Deaton found that means for household consumption were 22 percent lower in surveys than in national accounts and means of income from surveys were 43 percent lower than means from national accounts.

We also find that the size of the gaps varies systematically along income levels of countries, with the discrepancy being largest for middle-income countries. In contrast to previous assessments of gaps, we find relatively small differences between income and consumption estimates across all countries. Overall, the gaps across NAS and HHS for income and consumption measures are similar, which contradicts Ravallion's (2003) suggestion that the gap is mainly due to underreporting of incomes in surveys. In contrast to Deaton (2005) and Karshenas (2003), we show that the gap is narrowing as countries get richer for both income and consumption measures, possibly reflecting better integration of NAS and HHS data in high-income countries due to improved efforts to align survey and national accounts in recent years. We also find that growth rates from national accounts are higher than from surveys, particularly in middle-income countries, in line with the economic gradient in the gaps in levels.

We illustrate potential implications of these observed gaps for the measurement of global poverty and inequality. The implications depend on assumptions about the origin of the gaps and the corresponding adjustments made to the measures. We discuss implications of assuming that the gap is fully due to errors

in survey data and that these errors are distribution neutral (proportionally uniform across the income distribution). This is similar to scaling up survey means to national accounts means, in line with the methods applied by Bhalla (2002), Sala-i-Martin (2006) and Pinkovskiy and Sala-i-Martin (2014, 2016). Such adjustments generally show that both global poverty and inequality are lower and falling faster than measures based on surveys only.

We also examine a more nuanced explanation that assumes the gap in means is due to measurement (and definitional) differences across surveys and national accounts, but also with surveys disproportionately failing to fully capture the consumption and incomes of the richest households. We consider an adjustment similar to the approach proposed by Lakner and Milanovic (2016) and Chandy and Siedel (2017a, 2017b). We find that such adjustments have little effect on global poverty measures, but substantially revise upwards both global and national inequality, and significantly change our understanding of the distributional nature of growth and shared prosperity (i.e. consumption or income growth of the bottom 40 percent in each country). Notably, the relationship between observed levels of inequality and country income levels changes significantly, strengthening evidence of a Kuznets curve where inequality first rises and then falls with economic development (Kuznets, 1955).

The remainder of the paper proceeds as follows. Section 2 presents the data compiled for the analysis presented in this paper. Section 3 discusses reasons why gaps exist and presents our analysis of the differences between survey means and national accounts estimates (both in terms of levels and change over time) based on our newly compiled data. Section 4 considers two adjustments to survey data to reconcile these differences and describes the implications of each for global poverty and inequality measurement; the final section offers some concluding thoughts.

2. Data

2.1. *National Accounts*

Our main comparison between survey and national accounts data focuses on comparing the survey household consumption aggregate with the component of national accounts that corresponds to household expenditure, known as household final consumption expenditure (HFCE) and established in the 1993 System of National Accounts.[3] While per capita HFCE is the variable that conceptually most closely corresponds to the measures from household surveys (see Ravallion, 2003; Deaton, 2005; Anand and Segal, 2008 for example), we also compare survey estimates to per capita gross domestic product (GDP) estimates, for two reasons. First, GDP is a more frequently cited indicator of economic development and may be measured with less noise than HFCE, which is sometimes measured as a residual in the national accounts process. Second, this allows for comparison with literature that has compared survey means with GDP (Bourguignon and Morrison, 2002; Pinkovskiy and Sala-i-Martin, 2014, 2016).

[3]The definition of HFCE is broader than what is typically included in household surveys, encompassing items like spending of non-profit entities such as religious groups, NGOs and foundations. Previous literature (e.g. Ravallion, 2003; Deaton, 2005) has referred to HFCE as private consumption expenditure (PCE), defined under earlier systems of national accounts.

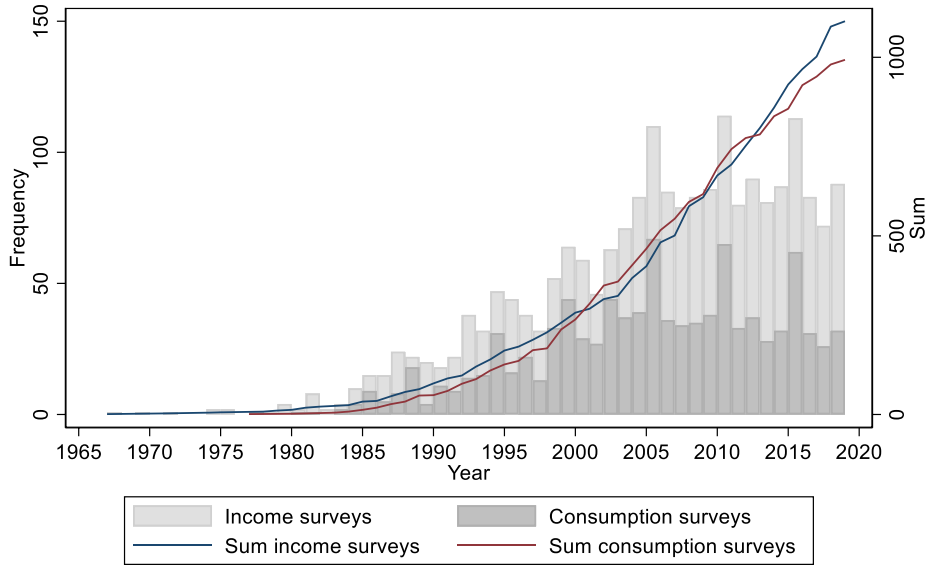


Figure 1. Available Consumption and Income Surveys Over Time

Notes: The bars show the number of surveys available for each year, separated by income (light gray) and consumption (dark gray), measured on the left-hand axis. The lines show the running sum of surveys over time, measured on the right-hand axis. Surveys that took place more than one calendar year are shown in the year in which data collection started. [Colour figure can be viewed at wileyonlinelibrary.com]

We extract national accounts data from the World Development Indicators (WDI) database, for both HFCE and GDP, using the series expressed both in current local currency units and in constant dollars. WDI's data is a compilation of World Bank and OECD national accounts data sets, obtained from official national sources. The per capita estimates are derived using the mid-year population estimates from the World Bank population series data.[4]

2.2. *Household Surveys*

To assess the gap between surveys and national accounts, we compile a data set of 2,095 national household survey means for 166 countries from 1967 until 2019, together covering countries that account for 97 percent of the world population in 2017. The distribution of surveys by type and over time is illustrated in Figure 1. The vast majority of the surveys come from PovcalNet, the World Bank's database for monitoring of global poverty (see Ferreira *et al.*, 2016 for a description of data sources and methods used). The database contains income or

[4]From World Development Indicators (WDI), we use the following series for national accounts data: Household final consumption expenditure (current LCU) [NE.CON.PRVT.CN]; Household final consumption expenditure (constant LCU) [NE.CON.PRVT.KN]; Household final consumption expenditure (constant 2010 US$) [NE.CON.PRVT.KD]; GDP (current LCU) [NY.GDP.MKTP.CN]; GDP (constant LCU) [NY.GDP.MKTP.KN]; GDP (constant 2010 US$) [NY.GDP.MKTP.KD]; and we use Population, total [SP.POP.TOTL] to construct per capita measures where needed. In the levels analysis we make use of the current LCU data, while in analysis of growth rates we use the constant series.

consumption distributions from nationally representative household surveys typically carried out or supervised by national statistical offices or international agencies, used for national and international poverty monitoring.

For most high-income countries, the survey data available in PovcalNet are for income (rather than consumption), originating from the Luxembourg Income Study and the European Union Statistics on Income and Living Conditions (EU-SILC). To ensure better coverage of consumption surveys from high-income countries in our sample, we supplement with data from other sources. For European countries, we derive consumption means from Eurostat's compilation of Household Budget Surveys. Eurostat publishes consumption means from household budget surveys, which are available in five-year intervals between 1988 and 2015. The Eurostat compilation provides consumption data for all 28 EU Member States and also for Montenegro, the Former Yugoslav of Republic of Macedonia, Turkey and Norway.[5] For the United States, we supplement our database with annual data from 1981 to 2015 from the Consumer Expenditure Survey.[6]

To facilitate comparisons with per capita national accounts data, we express all survey values (consumption and income) in annualized per capita basis. In cases where household survey consumption data are only available at the household level or in per adult equivalent terms, we adjust them to express in per capita terms. By converting all survey means to per capita terms, we align survey means with national accounts data.

Our comparisons of the gaps in levels in Section 3 are based on converting all data to local currency units. To accomplish this, we combine per capita mean consumption or income (expressed in 2011 PPP values) from each national survey with (i) inflation measures, (ii) currency conversion rates (in the case of currency devaluations and change of national currency) and (iii) the PPP values from PovcalNet that were used to convert the national survey data into 2011 PPP US dollars.[7] For the analysis in Section 4, where we examine how potential adjustments to account for the gaps between household surveys and national accounts affects global poverty and inequality measurement, we extract detailed information on the Lorenz curves for each welfare vector from PovcalNet. We are then able to construct national and global distributions of consumption and income from 1981 to 2017 that closely align with the World Bank's official estimates.[8]

[5]We obtain the Eurostat consumption data from the Eurostat Household Budget Survey Database available at https://ec.europa.eu/eurostat/web/household-budget-surveys/database (accessed on November 5, 2020). Eurostat provides survey means per adult equivalent in constant currency. We use available data about household structure from the corresponding surveys, as well as deflators and exchange rates to recover per capita means in current local currency units. With this data we can also recover estimates in 2011 PPPs, using the methods applied in PovcalNet.

[6]https://www.bls.gov/cex/, accessed on March 28, 2017.

[7]The data from PovcalNet used for this version of the paper was accessed March 28, 2021. The ancillary data on deflators and exchange dates is available at http://iresearch.worldbank.org/PovcalNet/data.aspx. PovcalNet has data for 168 economies, of which Argentina and Taiwan are not included in our compilation due to partial national geographic representativeness of the surveys.

[8]Following an approach similar to that taken by Lakner *et al*. (2019), we use the ungroup command included in the DASP Stata Package (Abdelkrim and Duclos, 2007) to generate a national distribution of 10,000 points for each reference year, based on Lorenz curves from PovcalNet. The resulting estimates of poverty and inequality are within 1 percentage point of direct PovcalNet estimates based on microdata in more than 95 percent of the cases.

For comparisons of growth as estimated by national accounts and surveys, there is the additional concern that heterogeneity in the survey data over time can potentially create noisy growth estimates. To partially address this concern, consumption and income surveys are considered separate series for each country, so that we do not calculate growth rates from an income survey in one year and a consumption survey in another. But there remain many ways in which growth estimates of consumption surveys (or income surveys) can be problematic due to changes in survey methods and practices. Jolliffe (2001), Beegle *et al*. (2012), Jolliffe *et al*. (2014) all document how changing instrument design over time within a country can change measured consumption, which will appear as growth (or contraction) over time independent of whether national wellbeing has in fact changed. As survey methods and questionnaires change over time within a country, PovcalNet typically does not post-adjust data from previous years for that country. Metadata on survey comparability suggests revisions in survey methods are widespread and create a break in comparability for two-thirds of the 164 countries that have data over time.[9] As further evidence of this problem, the World Bank is only able to produce comparable data for mean household survey growth required for the shared prosperity measure (requiring comparable data over a five-year period) for 88 of the 168 economies in its database.[10]

For the assessment of growth rates, where comparability in the measures matters significantly, we create an ancillary analysis file based on the World Bank's *Global Database of Shared Prosperity*. This database is designed with the objective of assessing growth as measured by comparable surveys within countries, over 3 to 7-year periods.[11] We use both the historical database which contains 467 spells for 121 countries, of which 88 countries were published in the most recent version in March 2021.[12] Since this database overlaps considerably with the PovcalNet database, we do not add these observations to our main data set, but rather create a complementary sample based on more recent, harmonized and comparable data with medium spell lengths useful for understanding how growth rates in surveys differ from those of national accounts. Results for this data set are presented as robustness checks in Section 4.

2.3. *Putting it All Together*

Of the 2,095 household survey estimates we compile; we match 2,082 to GDP per capita data and 2,020 to HFCE for the same country and year.[13] This sample is

[9]Criteria and data on survey estimate comparability are described in Atamanov *et al*. (2019). The comparability metadata is available in the World Bank's GitHub Repository for PovcalNet: https://raw.githubusercontent.com/worldbank/povcalnet/master/metadata/povcalnet_metadata.csv.

[10]Global Database of Shared Prosperity and Median Income/Consumption, circa 2013–2018, as of March 20, 2021: https://www.worldbank.org/en/topic/poverty/brief/global-database-of-shared-prosperity. A limitation of household surveys is that they are conducted with uneven frequency and with low consistency in methodology and implementation over time. Many countries lack surveys for at least five years or longer (Serajuddin *et al*., 2015).

[11]Global Database of Shared Prosperity, circa 2013–2018, available at http://www.worldbank.org/en/topic/poverty/brief/global-database-of-shared-prosperity (accessed May 8, 2021).

[12]See Yang and Nguyen (2021) for a description of the data and methodology. Access to the data is available at: https://www.worldbank.org/en/topic/poverty/brief/global-database-of-shared-prosperity.

[13]When a household survey spans more than one calendar year, we construct a weighted national accounts aggregate where the weights correspond to the distribution of months of survey fieldwork across the years. This is consistent with the practice of dealing with surveys in PovcalNet that run over multiple years, as described in Chen and Ravallion (2010).

substantially larger and more recent than those included in past studies. Deaton (2005) presented results from 557 surveys for 127 countries between 1979 and 2000. Ravallion (2003) used a smaller sample of survey means from 90 countries, and a panel of 142 growth spells for 60 countries. We also have substantially better coverage of High-Income countries, than Ravallion (2003), Karshenas (2003) and Deaton (2005).

We generate growth spells with annualized growth rates by matching all observations for each country and type (income/consumption), which gives more than 15,000 possible spells (including overlapping spells) and 1,881 consecutive (non-overlapping) spells. In contrast, Ravallion (2003) generates 142 spells between successive household surveys for 60 countries in the 1980s and 1990s. Our practice of matching all observations with each other within each country-type panel maximizes our potential power in estimation, but we also conduct analyses of spells that are non-overlapping, and with limited durations (e.g., spells of 3 years or less, spells shorter than 5 years).[14]

In all of our analyses, we treat the country as the unit of analysis and therefore weight each country equally. Where countries have varying number of observations of matched survey and national accounts data, each observation is weighted as the inverse of the number of observations for each country, so that the total weight given to each country, in each subsample, sums to one. Weights are re-calculated for each subsample. For example, if a country has two surveys – one consumption and the other income, each survey is given a weight of 0.5 in the pooled sample, while they each get a weight of one for the analysis when gaps for consumption are estimated separately from the gaps for income. This ensures that our analysis does not assign more weight to countries simply because they have more household surveys available, which is what would occur in an unweighted analysis. Our decision to treat the country as the unit of analysis is based on our interpretation of this literature as being fundamentally about the performance of country-level statistical systems, not people. Nonetheless, we do provide, supplemental tables with population-weighted and unweighted gaps in both levels and growth rates between surveys and national accounts as this can help explain how these gaps affect aggregate measures such as global changes in living standards or poverty. Comparable reviews, such as Deaton (2005) and Ravallion (2003) have used other weighting schemes than our preferred method. Supplemental tables with alternative weighting are available in Appendix 2.

3. Assessing the Gaps

There are many reasons why there are gaps between survey and national accounts estimates, but they largely fall into three categories—(i) measurement error in surveys, (ii) measurement error in national accounts, and (iii)

[14]We exclude 1 percent (top 0.5 and bottom 0.5) of estimated gaps in our main analyses of both levels and growth. Even after careful review of deflator and currency conversion issues, a few outlier observations remain, which are not suitable for inclusion in the main analyses as they reflect unrealistic levels of growth, and they reflect discrepancies likely due to currency, deflator or other errors.

conceptual differences in what each are designed to measure, including differences in their primary objectives. The evidence of measurement error in surveys is vast and indicative of substantial levels of noise in estimated levels of consumption and income. Examples demonstrating the sensitivity of estimated consumption to changes in questionnaire design or fieldwork protocols include Beegle *et al*. (2012), Browning *et al*. (2014), Engle-Stone *et al*. (2017), Jolliffe (2001), Pradhan (2009), Schündeln (2018), and Winter (2003). Most of these papers find a downward bias in estimated mean consumption compared to a benchmark. Consumption by logic (and instrument/questionnaire design) has a lower positive, nonzero bound, but has no symmetric upper bound. In part for this reason, measurement error in consumption, whether downward biased or mean preserving, tends to reduce the positive skewness in the distribution. Korinek *et al*. (2006) find higher non-response rates among the rich. Similarly, Banerjee and Piketty (2005) find that a substantial part of the HHS–NAS gap in India can be attributed to missing top incomes from surveys. Farfán *et al*. (2017) find significant underreporting of food-away-from home which would have the effect of compressing the positive tail of the distribution. Gibson *et al*. (2015) find that reporting error is negatively correlated with true consumption (accounting for which would require changing the distribution). The weight of the evidence is that measurement error in surveys is not distribution neutral and typically biases downward both mean consumption and the density of the upper tail of the distribution.

Within the limited literature of reconciling household surveys and national accounts there is sometimes a presumption that national accounts are the benchmark for comparison (i.e. they are free of measurement error). This inclination towards national accounts estimates over household surveys has been somewhat heightened recently by the assertion of Pinkovskiy and Sala-i-Martin (2016) that GDP per capita is better correlated than survey data with data on nighttime lights and non-monetary development outcomes, such as schooling and health status. Yet since the founding of national accounts in the 1940s, there has been an acknowledgement of measurement error in national accounts by those working in this field. As one of the seminal papers on this, Stone *et al*. (1942) introduced the idea of using balancing tables (balancing estimates from differing approaches) as a tool to reduce measurement error in national accounts estimates. More recently, Aruoba *et al*. (2016) uses this approach to estimate measurement error in US GDP estimates. Charmes (2012) discusses potential sources of measurement error in national accounts by providing an overview of how it struggles to capture informal economic activities. One specific type of informality, illegal activities are particularly challenging to measure in national accounts, and sometimes more readily picked up in household surveys. Buddenberg and Byrd (2006) note that in Afghanistan, national accounts are presumed to underestimate the economic value of the illegal drug industry, resulting in a potential downward bias of about one third in national income estimates. In contrast, they note that self-reports of poppy production by farmers in the national household survey for Afghanistan are substantial and do not appear to suffer from significant nonresponse problems. Despite these measurement concerns, it is useful to note that unlike household consumption

and income surveys, the global community has invested significant efforts in research, training and in general human capital development to produce reasonably harmonized measures of national income.

A final point to make on the observed gaps between household surveys and national accounts is to recognize that the objectives of both instruments differ. Likely of greatest relevance, household surveys are meant to measure the distribution of wellbeing of people (along many dimensions and frequently with a greater emphasis on the less well off), while national accounts are focused on measuring aggregate income and productivity (not the distribution of wellbeing). There is a long tradition of critiquing national accounts as a measure of wellbeing. Stiglitz *et al*. (2009) summarize many of these points noting that national income does not account for within-country distribution of income, is not monotonically increasing in wellbeing (e.g. traffic jams consume fuel, increasing national income but do not improve wellbeing), nor does it capture certain types of activities that contribute to wellbeing (e.g. unpaid household labor).

The overall objective of the System of National Accounts (SNA) is to produce an aggregate statistic. Deaton (2005) notes that SNA data tend to include larger transactions with greater probability than smaller transactions, and that to some extent this is intentional. The SNA training instructions specify that greater effort should be directed at larger transactions. Deaton cites OECD (2002, p. 179) where the SNA training instructions with respect to valuing home-production state that the time expended to collect this information should only be expended if the amount produced is sufficiently large with respect to the total supply in the nation of that particular item. In contrast, household budget surveys, and living standards surveys tend to focus on home production and more generally, include smaller transactions with greater probability than larger ones. The primary objectives of these surveys are typically to rank households, identify the poor, and measure consumption patterns. Conversely, estimates of consumption derived from these data frequently exclude large, once-in-a-lifetime, expenditures such as weddings and funerals as they tend to distort rankings of individuals, if not properly annuitized over the lifespan of the individual (Deaton and Zaidi, 2002).

A related concern with GDP is that despite significant efforts to establish international standards for the compilation of national accounts, guided by the UN Statistical Division's System of National Accounts (SNA), there remain substantial heterogeneity in methods and standards across countries. This heterogeneity has recently been particularly pronounced in Sub-Saharan Africa, where several countries have revised their national accounts estimates to incorporate new methods and data sources in efforts to better capture more economic sectors and include emerging economic activities that were not previously captured. These methodological revisions to national accounts can produce significant breaks in comparability. For example, in 2017, Senegal's GDP increased by 29 percent when it changed its base year from 1999 to 2014. In 2014, Nigeria's GDP nearly doubled when it rebased from 1990 to 2010 (Angrist *et al*., 2021). Similar revisions of more than 20 percent occurred in Ghana, Kenya, Senegal, and Zimbabwe (see Kouame *et al*., 2019). In contrast to the typical approach with household surveys, the common practice is to revise the entire national accounts series, improving comparability over time within each country.

Of potentially more consequence to cross-country comparability of GDP is the shifting of profits from one country to another by multi-national corporations. Paul Krugman coined the term *leprechaun economics* in 2015, when Ireland reported a 26 percent increase in GDP, which was mostly the result of an accounting maneuver by Apple to shift assets to Ireland on paper. This foreign direct investment in Ireland did not improve the wellbeing of the Irish in any meaningful way, nor did it increase economic activity, but was simply a tax-avoidance measure that resulted in distorting economic growth as measured by GDP. Damgaard *et al*. (2019) estimate that up to 40 percent of global foreign direct investment is phantom investment with no tangible links to the local economy.

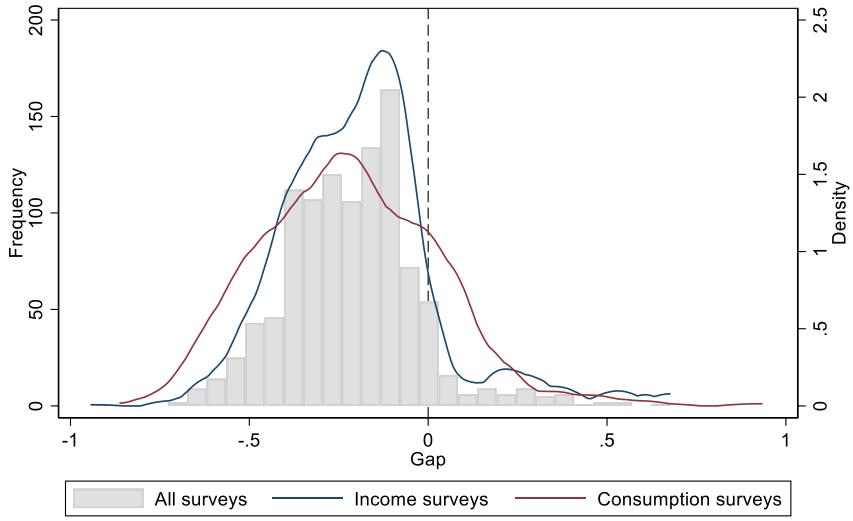
3.1. *HHS–NAS Gaps, Levels*

With these being some key reasons for the observed gaps between national accounts and household surveys, we now turn to empirically estimating their magnitude. Following the approach of Ravallion (2003) and Deaton (2005), we calculate a set of ratios of the survey to national accounts means. By subtracting one from the ratio, we get a measure of the proportional gap, between the HHS mean and the NAS mean. A negative (positive) gap suggests the survey mean is lower (higher) than the corresponding national accounts mean. For example, a ratio of 0.9, gives a gap measure of −0.1, suggesting that the survey mean is 10 percent lower than the national accounts mean.

The distribution of the gap in our sample of matched survey and national account means is illustrated in Figure 2. Both the unweighted histogram and the country-weighted density functions show that, on average, the gap is well below zero, indicating that survey means on average are lower than national accounts means. The gap is substantially larger with regards to GDP than with HFCE, moreover consumption shows larger variation than income, as reflected in the wider density functions seen for consumption surveys. We estimate the average gap for various subsamples and calculate standard errors clustered at the country level. Specifically, we estimate an OLS regression of the observed gap on a constant: $g_i^n = \alpha + \varepsilon$ for various subsamples and weighting schemes, where α gives us the mean gap. We cluster the variance estimates at the country level allowing for errors to be correlated across observations within countries.

Table 1 shows the value of the gaps for various sub-samples, by region and survey type. Overall, our estimates suggest that survey means are 20 percent lower than national accounts means for HFCE. There is no statistically significant difference in the gap observed for consumption (22 percent) and income (20 percent). Notably, we find the gaps are substantially larger than the corresponding estimates from Ravallion (2003) and Deaton (2005). Deaton found a gap of 14 percent for consumption and 10 percent for income, while Ravallion estimated a gap of 33 percent for income and 7 percent for consumption, with the difference between income and consumption measures being statistically significant. Karshenas (2003) estimates the gap using mean difference in the logs and finds a difference of approximately 5 percent. While our estimates differ substantially from those of Deaton and Ravallion, we find this difference

Panel A: Survey mean vs HFCE mean



Panel B: Survey mean vs GDP mean

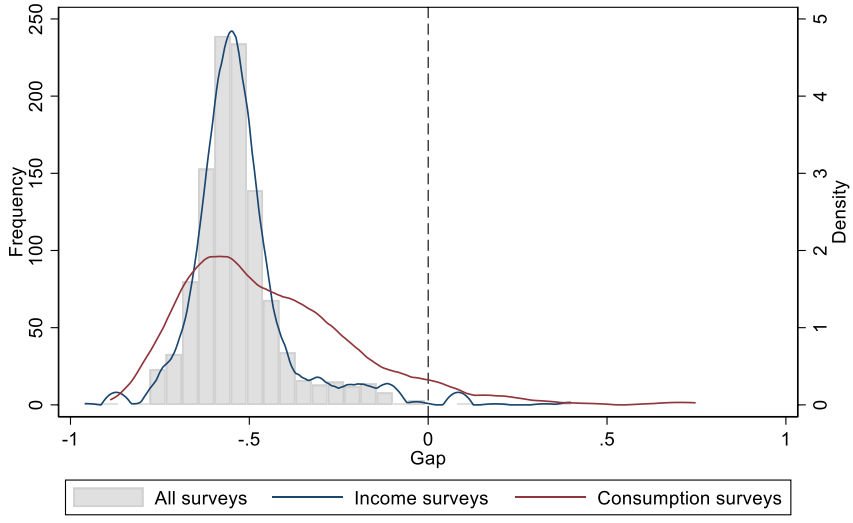


Figure 2. Distribution of Gap of Survey to National Accounts Consumption.

Notes: The charts show the distribution of the estimated gap between household survey mean and national accounts means. The gap is estimated as the proportional difference (per cent) of survey means with respect to national accounts means. It is estimated as the ratio of per capita survey mean over the per capita national accounts mean, minus one. A negative (positive) gap estimate indicates that surveys is (larger) than national account means. The histograms are unweighted; the density functions are weighted to give each country equal weight, using the same approach as in Table 1. Outlier observations above 1 are excluded (one observation for GDP and three observations for HFCE). [Colour figure can be viewed at wileyonlinelibrary.com]

disappears when we compare our estimates over similar time periods. The Deaton analysis includes data from 1979 to 2000, while the Ravallion analysis covers 1985–1998. Our estimated overall difference between household survey means (pooling consumption and income) and HFCE for the 1980s and 1990s are 15 percent. These estimates are similar to Ravallion's 17 percent (95% confidence interval of approximately 9 to 25 percent) and Deaton's 12 percent (95% confidence interval of approximately 6 to 18 percent). One inference to be drawn from this analysis is that our estimated gaps are substantially larger from those estimated separately by Deaton and Ravallion, but this is not due to differences in methods or assumptions, but due primarily to our data including more recent observations. We estimate the overall average gap during the 2010 decade to be 75 percent greater in absolute magnitude relative to the average gap during the 1990s. For a more detailed comparison of our findings with those of Deaton (2005) and Ravallion (2003), see Appendix 1.

Across all sub-samples by geography and income groups, the gap is negative and statistically significant (Table 1). In Panel A of Table 1, we compare all survey measures (overall, consumption, income) with HFCE. There is a marked pattern of the ratio and country income levels, with the gap being largest among middle-income countries (23 percent for lower middle-income countries and 33 percent for upper middle-income countries). The gap is much smaller for low- and high-income countries, but still statistically significantly different from zero. The pattern across levels of economic development is shown in greater detail in Figure 3, which plots all gaps (and highlights the most recent observations) against economic development, as measured by GNI per capita. The relationship between the size of the gap and level of income forms a clear u-shape, indicating that, on average, the gap increases (gets more negative) from low income to middle-income country range, but then is diminishes for high-income countries. The lines show the results of locally weighted regressions of the latest observation for each country.

The pattern in the Figure 3 differs from the findings by Deaton (2005), Ravallion (2003) and Karshenas (2003). Deaton and Karshenas both find a negative relationship between the gap and country income levels, with the gap being largest among the richest countries. In contrast, our sample shows that the gap is smaller among high-income countries than among middle income countries. An important part of the explanation for this is that compared to previous studies, our database contains more surveys overall from richer countries, and also more consumption surveys for richer countries and more income surveys for poorer countries. Moreover, we observe a very similar pattern for income surveys and consumption surveys, which Deaton and Ravallion did not. We believe this difference originates from income surveys being more heavily concentrated in richer countries in the Deaton and Ravallion samples, where they observed a larger gap, while consumption surveys were concentrated in poorer countries, where their gap was smaller. This pattern was particularly strong in the case of Deaton who focused the analysis of rich countries on the UK and the US, two high-income countries where the gap is particularly large. We also find relatively larger gaps in the UK and US, but these appear somewhat unique cases among rich countries.

A further possible reason we observe a smaller and narrowing gap among richer countries than previous studies may be linked to evolving practices and

TABLE 1
GAPS IN SURVEY MEAN AND NATIONAL ACCOUNTS MEANS, ACROSS REGIONS AND INCOME GROUPS

Group	All Surveys			Consumption Surveys			Income Surveys		
	Obs.	Gap	S.E.	Obs.	Gap	S.E.	Obs.	Gap	S.E.
Panel A. Gap in levels between survey mean and HFCE mean									
All	2020 (156)	−0.202***	0.018	947 (132)	−0.221***	0.019	1073 (72)	−0.200***	0.027
1980s	118 (57)	−0.149***	0.030	52 (36)	−0.154***	0.038	66 (30)	−0.166***	0.037
1990s	358 (113)	−0.145***	0.037	196 (92)	−0.161***	0.044	162 (54)	−0.167***	0.034
2000s	747 (140)	−0.244***	0.017	368 (116)	−0.262***	0.019	379 (58)	−0.242***	0.021
2010s	792 (142)	−0.254***	0.018	331 (116)	−0.287***	0.019	461 (61)	−0.228***	0.024
Low income	198 (50)	−0.129***	0.037	189 (47)	−0.137***	0.036	9 (7)	−0.055	0.129
Lower middle income	503 (87)	−0.232***	0.026	338 (74)	−0.266***	0.025	165 (23)	−0.117*	0.068
Upper middle income	524 (55)	−0.326***	0.025	238 (42)	−0.352***	0.033	286 (27)	−0.337***	0.030
High income	768 (42)	−0.216***	0.023	168 (30)	−0.272***	0.026	600 (41)	−0.209***	0.024
East Asia & Pacific	160 (18)	−0.150***	0.046	117 (14)	−0.170**	0.058	43 (6)	−0.127*	0.054
Europe & Central Asia	1072 (49)	−0.253***	0.032	484 (47)	−0.275***	0.033	588 (38)	−0.229***	0.036
Latin America & Caribbean	392 (21)	−0.199***	0.055	29 (6)	−0.316***	0.074	363 (20)	−0.192***	0.059
Middle East & North Africa	91 (14)	−0.160***	0.048	65 (12)	−0.193***	0.039	26 (3)	−0.034	0.166
North America	81 (2)	−0.189	0.119	32 (1)	−0.335	0.000	49 (2)	−0.175	0.105
South Asia	46 (7)	−0.245**	0.080	45 (7)	−0.244**	0.081			
Sub-Saharan Africa	178 (45)	−0.174***	0.034	175 (45)	−0.170***	0.035	3 (2)	−0.276	0.064
Panel B. Gap in level between survey mean and GDP mean									
All	2082 (165)	−0.465***	0.015	990 (138)	−0.462***	0.018	1092 (75)	−0.516***	0.017
1980s	122 (61)	−0.459***	0.027	53 (37)	−0.439***	0.041	69 (33)	−0.506***	0.022
1990s	385 (122)	−0.422***	0.027	208 (97)	−0.414***	0.033	177 (58)	−0.484***	0.025
2000s	764 (150)	−0.484***	0.016	385 (126)	−0.487***	0.019	379 (58)	−0.545***	0.015
2010s	806 (150)	−0.508***	0.015	344 (123)	−0.518***	0.017	462 (62)	−0.546***	0.014
Low income	219 (55)	−0.304***	0.033	209 (52)	−0.299***	0.035	10 (8)	−0.224*	0.116
Lower middle income	534 (95)	−0.485***	0.018	356 (81)	−0.505***	0.020	178 (25)	−0.412***	0.036
Upper middle income	532 (60)	−0.578***	0.019	242 (45)	−0.590***	0.025	290 (29)	−0.593***	0.017
High income	768 (42)	−0.576***	0.009	168 (30)	−0.606***	0.013	600 (41)	−0.571***	0.009
East Asia & Pacific	175 (23)	−0.431***	0.040	132 (19)	−0.415***	0.048	43 (6)	−0.519***	0.006

(Continues)

TABLE 1 (CONTINUED)

Group	All Surveys			Consumption Surveys			Income Surveys		
	Obs.	Gap	S.E.	Obs.	Gap	S.E.	Obs.	Gap	S.E.
Europe & Central Asia	1076 (49)	−0.542***	0.022	486 (47)	−0.549***	0.024	590 (38)	−0.548***	0.025
Latin America & Caribbean	410 (24)	−0.463***	0.034	30 (6)	−0.432***	0.083	380 (23)	−0.464***	0.036
Middle East & North Africa	96 (15)	−0.460***	0.035	70 (13)	−0.446***	0.038	26 (3)	−0.546***	0.026
North America	81 (2)	−0.511**	0.034	32 (1)	−0.564	0.000	49 (2)	−0.501**	0.024
South Asia	48 (7)	−0.535***	0.053	47 (7)	−0.534***	0.054			
Sub−Saharan Africa	196 (45)	−0.386***	0.037	193 (45)	−0.385***	0.037	3 (2)	−0.586**	0.041

Notes: The gap is estimated as the proportional difference (per cent) of survey means with respect to national accounts means. It is estimated as the ratio of per capita survey mean over the per capita national accounts mean, minus one. A negative (positive) gap estimate indicates that surveys means within the sample are smaller (larger) than national account means. Number of countries in each group given in parenthesis. Robust standard errors clustered at the country level (based on sandwich estimator). Statistical significance from zero, denoted by $^{*}P \leq 0.1$, $^{**}P \leq 0.05$, $^{***}P \leq 0.01$. Each country is given equal weight within each sub-sample/group estimate, regardless of the number of observations for the country. Additional estimates with alternatives weights are available in Appendix 2. The column labelled "All surveys" pools consumption and income surveys. Due to weights and the fact that not all countries having both consumption and income surveys, estimates in "All surveys" does in some cases not fall within range of the estimates for the samples separated by survey type.

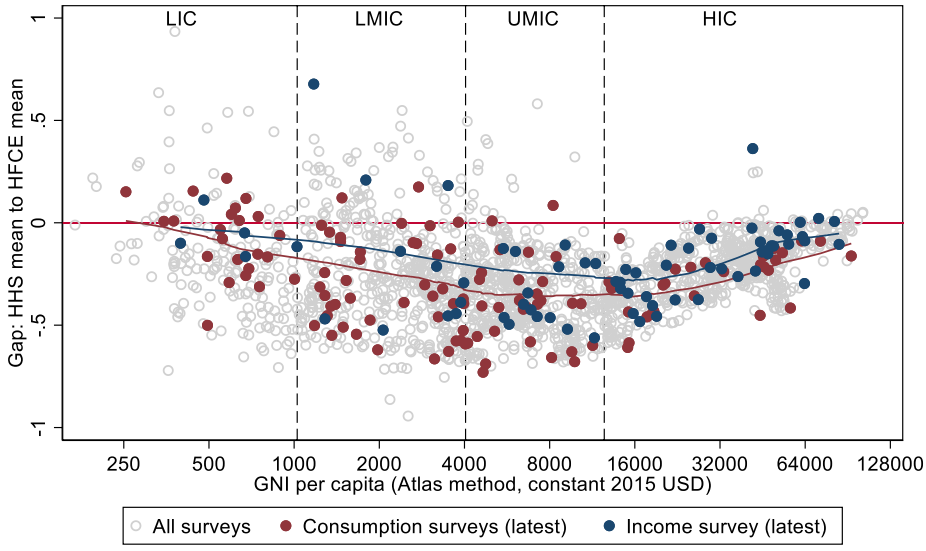


Figure 3. Survey-national Accounts Gap and Level of Development

Notes: The gap is estimated as the proportional difference (per cent) of survey means with respect to national accounts means with negative estimates meaning that household survey means are smaller than corresponding measures for national accounts. The lowess lines (locally weighted scatterplot smoothing) are based on a non-parametric regression with bandwidth of 0.8. The vertical bands demarcate the cutoffs for the World Bank's income classifications, expressed in 2015 USD, deflated using the Atlas method based on the SDR deflator. [Colour figure can be viewed at wileyonlinelibrary.com]

consolidation of data sources in these countries. Recently, it has been more common in rich countries to integrate administrative data, such as tax records, into survey data estimates. For example, the EU-SILC surveys, a major source for our income survey data for rich countries, explicitly allow countries to supplement income variables with administrative records, such as tax and payroll data (Jäntti *et al*., 2013). This practice helps align estimates, reduces underreporting in surveys, and contributes to a reduction in the gap between national accounts measures and survey measures. A caveat to this interpretation is that consumption surveys, which do not typically rely on any such administrative data, also show a similar narrowing of the gap among rich countries. In Panel B of Table 1, we also assess the gap between household surveys and GDP. Overall, survey means are on average 47 percent lower than GDP across countries. While this large discrepancy may not be surprising to national accounts and survey experts, it is worth noting that maybe the most commonly cited indicator summarizing living standards in a country—GDP per capita—and often described as "income per person," is on average twice the size of per capita household income or consumption from surveys. Interestingly, we do not see as much narrowing of the gap among the richest countries for GDP, as with HFCE. This results from a widening of the gap between HFCE and GDP over time. Upper-middle-income countries have the largest gap, on average, but it is not much larger than that of the high-income countries in our

sample. Additional estimates for gaps in levels with different weighting schemes are available in Appendix 2, Table A2.1.

The larger gaps between national accounts and surveys observed in middle-income countries may reflect the difficulty of measuring consumption in household surveys in the context of rapid growth and associated changes in consumption patterns. The experience of revisions to Vietnam's consumption survey in 2010 is pertinent to this (see Badiani *et al.*, 2013). From the early 1990s to 2010, a period of rapid economic growth in Vietnam, the national statistics office made few changes to the consumption survey instrument and methodology, resulting in a failure at capturing new durable consumption items, such as cellphone and computers. By 2010, the country's much greater affluence necessitated the use of a revised consumption questionnaire and survey methodology. Consumption measures from the old survey methodology, which reflected the consumption patterns in Vietnam in 1992, was just 78 percent of what was measured as the average household budget in 2010 using the new survey instrument and the new methodology for measuring household consumption. Household consumption in the surveys leading up to 2010 were falling further and further below the consumption captured in the revised survey instrument. The example of Vietnam gives some context as to how the surveys fail to capture consumption as countries get richer. At the same time, one should recognize the importance of maintaining consistent series of consumption data over time. One approach to combine consistency and revise methodology would be to generate several consumption aggregates chaining new and old survey methods, as commonly done when revising national accounts.

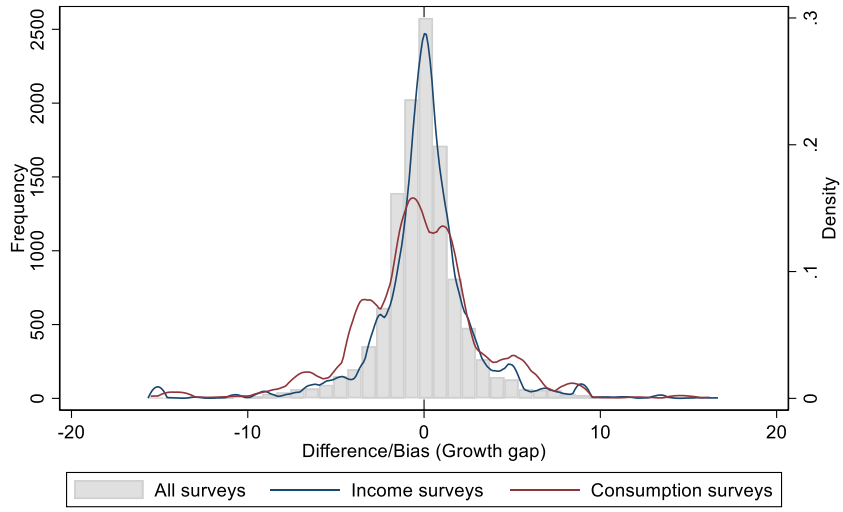
3.2. *HHS–NAS Gaps, Growth Rates*

We now turn to assessing how growth rates in national accounts differ from those of household surveys. To assess the relationship between growth rates in the per capita means of surveys and national accounts measures, we follow methods commonly used in evaluating forecasts of economic growth.[15]We take the difference in annualized growth rates of survey means and national accounts as our measure of the gap (or "error" in the growth forecast literature). A negative value indicates that survey means grew slower than the national accounts measures.

The distribution of the gap is presented in Figure 4, showing a large variation in differences in growth rates between the two sources. To check for systematic differences, we estimate these differences in growth rates by subgroups of countries, time periods and spell lengths, for each type of survey (income and consumption). We estimate subgroup averages by running a simple regression of the growth rate gap on a constant, analogous to the preceding assessment of differences in levels. The main results are presented in Table 2. Overall, growth rates in surveys are, on average, lower than in national accounts. However, the difference is not statistically significant for the overall sample which pools both income and consumption for all countries and time periods. For the subsamples

[15]The growth forecasting literature is concerned with assessing the precision and bias of forecasts in predicting actual growth rates (see for example, Holden and Peel, 1990; Artis and Marcellino, 2001). Similarly, we are interested in understanding the precision and bias of national accounts growth rates in predicting actual survey growth rates.

Panel A: HFCE growth vs household survey growth



Panel B: GDP growth vs household survey growth

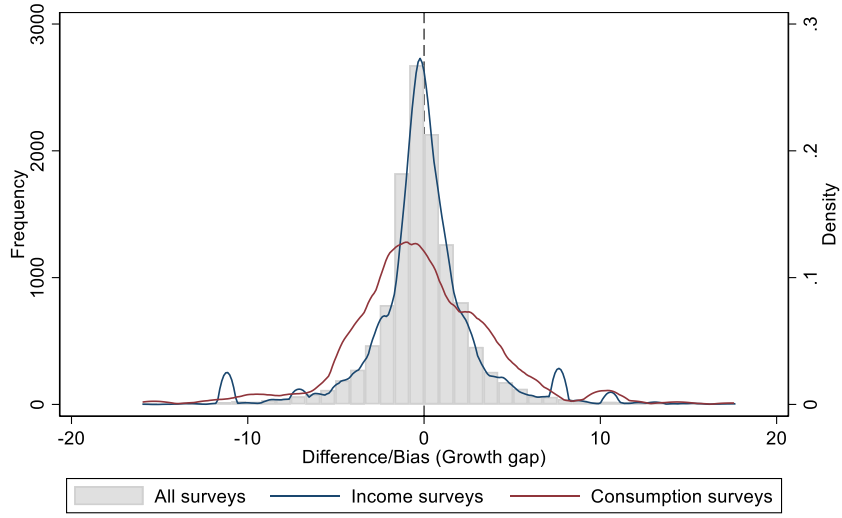


Figure 4. Gap in Growth Rates in National Accounts vs. Household Surveys.

Notes: The "growth gap" is estimated as the difference between annualized growth rate in surveys and the growth in national accounts. A negative difference (gap) means that household survey means grew slower than corresponding measures for national accounts. The lowess lines (locally weighted scatterplot smoothing) are based on a non-parametric regression with bandwidth of 0.8. The vertical bands demarcate the cutoffs for the World Bank's income classifications, expressed in 2015 USD, deflated using the Atlas method based on the SDR deflator. There is large variation in the gap and the distribution for consumption surveys is skewed to the left of zero, indicating that growth rates in consumption surveys typically is slower than in national accounts. Details are available in Table 2. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2
GROWTH IN SURVEYS VS NATIONAL ACCOUNTS—BIAS AND ERRORS

Group	All Surveys				Consumption Surveys				Income Surveys			
	Obs.	Bias	S.E.	Error	Obs.	Bias	S.E.	Error	Obs.	Bias	S.E.	Error
Panel A. Growth gap between survey mean and HFCE mean												
All spells	15062 (133)	−0.25	0.18	2.25	5068 (108)	−0.53***	0.18	2.35	9994 (61)	0.04	0.17	1.79
All 1 year spells	9588 (128)	−0.13	0.18	1.89	3064 (102)	−0.31*	0.18	1.96	6524 (59)	0.10	0.14	1.43
All 1 to 3 year spells	3310 (97)	−0.28	0.26	3.39	1103 (54)	−0.84***	0.26	4.17	2207 (60)	0.07	0.34	2.91
All 4 to 6 year spells	3066 (119)	−0.25	0.20	2.26	1133 (94)	−0.53***	0.20	2.43	1933 (60)	0.10	0.18	1.62
Comparable spells only	7577 (122)	−0.29	0.19	2.44	2322 (80)	−0.89***	0.19	2.98	5255 (59)	0.11	0.17	1.95
Comparable, bal, 1−3 year spells	2639 (88)	−0.01	0.29	3.24	820 (44)	−0.99***	0.29	4.15	1819 (58)	0.39	0.29	2.78
Comparable, bal, 4−6 year spells	2017 (88)	−0.44**	0.18	1.86	599 (44)	−1.28***	0.18	2.43	1418 (58)	0.03	0.18	1.54
1980s	74 (20)	1.12	0.90	3.08	22 (6)	−0.59	0.90	1.50	52 (14)	1.85	1.23	3.76
1990s	413 (71)	−1.57***	0.40	3.67	153 (50)	−2.06***	0.40	3.57	260 (39)	−0.95**	0.44	3.27
2000s	2010 (102)	−0.18	0.23	2.82	772 (62)	−0.65***	0.23	3.19	1238 (53)	0.59*	0.33	2.60
2010s	2418 (94)	−0.41*	0.21	2.10	722 (68)	−0.94***	0.21	2.11	1696 (56)	0.24	0.19	1.65
Low income	372 (39)	−0.41	0.64	3.59	361 (38)	−0.18	0.64	3.44	11 (2)	−4.77	4.92	5.24
Lower Middle Income	2253 (65)	−1.10***	0.39	3.21	1202 (53)	−1.66***	0.39	3.27	1051 (18)	−0.36	0.80	3.69
Upper Middle Income	3085 (45)	−0.32	0.23	2.41	964 (32)	−0.70***	0.23	2.58	2121 (26)	0.27	0.42	2.48
High Income	5872 (40)	0.12	0.19	1.61	811 (29)	−0.59***	0.19	1.79	5061 (38)	0.33*	0.17	1.50
East Asia & Pacific	1029 (12)	−0.39	0.63	2.37	856 (8)	−0.73	0.63	2.48	173 (5)	−0.11	0.79	2.10
Europe & Central Asia	8109 (47)	−0.20	0.17	1.93	2973 (45)	−0.69***	0.17	2.11	5136 (33)	0.34***	0.13	1.39
Latin America & Caribbean	4131 (19)	−0.30	0.42	2.33	112 (3)	−3.85***	0.42	4.15	4019 (19)	−0.21	0.43	2.30
Middle East & North Africa	244 (9)	−0.79**	0.34	2.14	196 (7)	−0.56	0.34	1.63	48 (2)	−1.57	1.35	3.93
North America	1114 (2)	−0.87**	0.37	0.94	496 (1)	−1.53***	0.37	1.57	618 (2)	−0.71***	0.21	0.80
South Asia	170 (6)	−0.29	0.58	1.94	170 (6)	−0.29	0.58	1.94				
Sub−Saharan Africa	265 (38)	−0.06	0.50	2.70	265 (38)	−0.06	0.50	2.70				
Panel B. Growth gap between survey mean and GDP mean												
All spells	16089 (155)	−0.19	0.19	2.45	5678 (127)	−0.43**	0.19	2.53	10411 (67)	−0.09	0.26	2.03
All 1 year spells	10139 (145)	−0.09	0.17	1.96	3398 (117)	−0.27	0.17	2.02	6741 (62)	0.06	0.20	1.58
All 1 to 3 year spells	3464 (109)	−0.12	0.26	3.53	1186 (65)	−0.59**	0.26	4.28	2278 (62)	0.21	0.25	2.87

(Continues)

TABLE 2 (CONTINUED)

Group	All Surveys				Consumption Surveys				Income Surveys			
	Obs.	Bias	S.E.	Error	Obs.	Bias	S.E.	Error	Obs.	Bias	S.E.	Error
All 4 to 6 year spells	3239 (135)	−0.41*	0.23	2.54	1243 (107)	−0.60**	0.23	2.73	1996 (65)	−0.19	0.24	1.83
Comparable spells only	8010 (136)	−0.40**	0.18	2.45	2639 (92)	−0.94***	0.18	2.86	5371 (62)	0.05	0.17	2.04
Comparable, bal, 1–3 year spells	2767 (95)	−0.11	0.28	3.37	892 (50)	−0.99***	0.28	4.09	1875 (60)	0.32	0.28	2.88
Comparable, bal, 4–6 year spells	2116 (95)	−0.64***	0.20	2.10	662 (50)	−1.47***	0.20	2.70	1454 (60)	−0.06	0.16	1.64
1980s	85 (23)	0.11	0.96	3.21	27 (8)	−2.35**	0.96	2.74	58 (15)	1.42	1.21	3.46
1990s	493 (84)	−1.28***	0.41	3.76	187 (60)	−1.55***	0.41	3.88	306 (42)	−1.02**	0.42	3.05
2000s	2097 (118)	−0.17	0.23	2.79	851 (77)	−0.58**	0.23	2.97	1246 (54)	0.71**	0.32	2.69
2010s	2481 (103)	−0.29	0.23	2.21	741 (76)	−0.79***	0.23	2.34	1740 (58)	0.01	0.22	1.80
Low income	474 (49)	−0.44	0.49	3.53	460 (48)	−0.34	0.49	3.50	14 (3)	−2.61	2.08	2.89
Lower Middle Income	2676 (77)	−0.67*	0.40	3.24	1363 (65)	−1.06***	0.40	3.37	1313 (18)	−0.45	0.76	3.23
Upper Middle Income	3135 (50)	−0.04	0.41	2.90	999 (35)	−0.27	0.41	2.87	2136 (28)	0.33	0.69	3.13
High Income	5955 (41)	0.11	0.18	1.54	817 (30)	−0.63***	0.18	1.75	5138 (39)	0.19	0.18	1.51
East Asia & Pacific	1129 (19)	−0.40	0.51	2.33	952 (15)	−0.62	0.51	2.29	177 (6)	−0.28	0.82	2.24
Europe & Central Asia	8501 (48)	−0.29	0.24	2.23	3301 (46)	−0.67***	0.24	2.43	5200 (33)	0.13	0.16	1.53
Latin America & Caribbean	4422 (23)	−0.03	0.68	2.98	133 (4)	−2.21***	0.68	4.54	4289 (22)	−0.04	0.71	2.80
Middle East & North Africa	338 (12)	−1.09***	0.32	1.69	212 (10)	−1.30***	0.32	1.80	126 (3)	−0.72	0.75	1.48
North America	1114 (2)	−0.56	0.41	0.73	496 (1)	−1.28***	0.41	1.34	618 (2)	−0.39	0.24	0.59
South Asia	175 (7)	0.13	0.96	2.35	175 (7)	0.13	0.96	2.35				
Sub−Saharan Africa	410 (44)	0.14	0.42	2.79	409 (44)	0.17	0.42	2.76				

Notes: Robust standard errors clustered at the country level (based on robust sandwich estimator). Statistical significance of difference from zero, denoted by $*P \leq 0.1$, $**P \leq 0.05$, $***P \leq 0.01$. Within each group (sample) each country is given equal weight regardless of number of country−year observations. Number of countries in each group given in parentheses. Comparable spells are calculated only between surveys within a country that are deemed comparable for the purposes of international poverty monitoring. Details about the comparability indicator is available in Atamanov *et al.* (2019). The subsamples labelled "balanced" ensures the composition of countries across spell lengths is constant and useful for understanding how spell length affect estimates without affecting the composition of countries.

of lower middle-income countries, as well as the 1990s, survey means grow more than one percentage point slower than HFCE in national accounts on average, a difference that is statistically different from 0 at the 95%-level. The gap in growth rates is also larger when looking only at longer, comparable spells for each country (with four-to-six-year spells).

The gaps in growth rates between surveys and national accounts are larger and more robust when looking at the subsamples consisting of only survey consumption means. Overall, we find that the average growth rate for consumption means in surveys in our full sample is 0.53 percentage points less than HFCE and 0.43 percentage points less than GDP growth. The difference is particularly large in middle-income countries. For income means, the differences in growth rates with respect to national accounts are smaller and mostly statistically insignificant. Notably, we observe a positive gap for high-income countries, where survey means grow faster than HFCE. This is despite the North America region (where both the US and Canada are also considered high income) showing income surveys grow more slowly than HFCE. Additional estimates for gaps in growth rates using different weighting schemes are available in Appendix 2, Table A2.2.

The general pattern of a larger gap in growth in middle-income countries is reasonably consistent with the findings regarding the gap in levels. The gap in levels is increasing across the middle-income countries, suggesting that surveys also grow slower than national accounts means over this segment. For high-income countries, the point estimate suggests that growth rates of consumption in surveys is higher than that in national accounts, which contributes to explaining the narrowing gap seen for high-income countries (Figure 3). The narrowing of the gaps as countries get richer could be due to the increased integration of survey data with national accounts in richer countries, noted earlier. Notably, in US and Canada, this integration between survey and national accounts measures is not practiced to the same extent, and a larger and widening gap is seen in these countries.

In addition to the direction of bias (or gaps), we are often interested in the precision (efficiency) of growth rates of national accounts as a predictor of survey growth rates, given that national accounts data are frequently used to interpolate and extrapolate household survey consumption or income to obtain estimates of poverty in years surveys are not conducted, as done by the World Bank. The mean absolute error and the root mean square errors presented in Table 2 gives an assessment of the precision, which is relevant for using national accounts in predicting survey growth. Again, there is a clear economic gradient with the precision increasing with income, but it should be noted that high-income countries typically experience lower growth rates and the gradient of relative errors may be smaller. The mean absolute error is large, more than 2 percentage points for all sub-groups except for North America and high-income countries, highlighting the large average differences in growth rates observed between HHS and NAS data sources.

Ravallion (2003) uses an alternative way to assess the degree of correspondence of growth rates, using a simple no-constant OLS regression of survey means on national account means. Table 3 presents results using this approach. The resulting coefficients on the national accounts mean have typically been used when extrapolating income and consumption distributions for the World Bank's poverty projections beyond the latest official reference year, such as

TABLE 3
GROWTH IN SURVEYS VS NATIONAL ACCOUNTS—NO−CONSTANT REGRESSION RESULTS

Group	All				Consumption				Income			
	Obs.	Coeff	S.E.	r^2	Obs.	Coeff	S.E.	Error	Obs.	Coeff	S.E.	Error
Panel A. No−constant OLS regression of survey mean and HFCE mean												
All spells	15062 (133)	0.74***	0.05	0.46	5068 (108)	0.73***	0.05	0.73	9994 (61)	0.89	0.13	0.89
All 1 year spells	9588 (128)	0.71***	0.06	0.71	3064 (102)	0.68***	0.06	0.68	6524 (59)	0.88	0.16	0.88
All 1 to 3 year spells	3310 (97)	0.75***	0.09	0.75	1103 (54)	0.73***	0.09	0.73	2207 (60)	0.87	0.20	0.87
All 4 to 6 year spells	3066 (119)	0.80***	0.06	0.80	1133 (94)	0.76***	0.06	0.76	1933 (60)	1.01	0.07	1.01
Comparable spells only	7588 (122)	0.75***	0.05	0.75	2325 (80)	0.67***	0.05	0.67	5263 (59)	1.05	0.08	1.05
Comparable, bal, 1–3 year spells	2460 (82)	0.74***	0.08	0.74	747 (40)	0.62***	0.08	0.62	1713 (54)	1.05	0.07	1.05
Comparable, bal, 4–6 year spells	1880 (83)	0.81***	0.05	0.81	558 (42)	0.71***	0.05	0.71	1322 (55)	1.01	0.07	1.01
1980s	93 (23)	0.62	0.28	0.62	24 (6)	0.45**	0.28	0.45	69 (18)	0.64	0.32	0.64
1990s	413 (71)	0.64***	0.10	0.64	153 (50)	0.64***	0.10	0.64	260 (39)	0.55***	0.14	0.55
2000s	2010 (102)	0.85***	0.05	0.85	772 (62)	0.78***	0.05	0.78	1238 (53)	1.10	0.09	1.10
2010s	2418 (94)	0.67***	0.08	0.67	722 (68)	0.56***	0.08	0.56	1696 (56)	1.01	0.17	1.01
Low income	372 (39)	0.64**	0.17	0.64	361 (38)	0.70*	0.17	0.70	11 (2)	−0.13**	0.49	−0.13
Lower Middle Income	2253 (65)	0.69***	0.07	0.69	1202 (53)	0.66***	0.07	0.66	1051 (18)	0.67	0.33	0.67
Upper Middle Income	3085 (45)	0.89**	0.05	0.89	964 (32)	0.79***	0.05	0.79	2121 (26)	1.18*	0.10	1.18
High Income	5872 (40)	0.78	0.20	0.78	811 (29)	0.67	0.20	0.67	5061 (38)	0.87	0.25	0.87
East Asia & Pacific	1029 (12)	0.78**	0.09	0.78	856 (8)	0.77**	0.09	0.77	173 (5)	0.70**	0.13	0.70
Europe & Central Asia	8109 (47)	0.87***	0.05	0.87	2973 (45)	0.81***	0.05	0.81	5136 (33)	1.14**	0.06	1.14
Latin America & Caribbean	4131 (19)	0.92	0.12	0.92	112 (3)	0.32***	0.12	0.32	4019 (19)	0.94	0.12	0.94
Middle East & North Africa	244 (9)	0.15***	0.21	0.15	196 (7)	0.44***	0.21	0.44	48 (2)	−0.16***	0.21	−0.16
North America	1114 (2)	0.55***	0.17	0.55	496 (1)	0.27***	0.17	0.27	618 (2)	0.63***	0.10	0.63
South Asia	170 (6)	0.68**	0.14	0.68	170 (6)	0.68**	0.14	0.68				
Sub-Saharan Africa	265 (38)	0.54***	0.14	0.54	265 (38)	0.54***	0.14	0.54				
Panel B. No-constant OLS regression of survey mean and GDP mean												
All spells	16089 (155)	0.79***	0.05	0.39	5678 (127)	0.75***	0.05	0.75	10411 (67)	0.94	0.08	0.94
All 1 year spells	10139 (145)	0.80***	0.06	0.80	3398 (117)	0.76***	0.06	0.76	6741 (62)	0.93	0.07	0.93
All 1 to 3 year spells	3464 (109)	0.76***	0.08	0.76	1186 (65)	0.67***	0.08	0.67	2278 (62)	1.02	0.08	1.02
All 4 to 6 year spells	3239 (135)	0.77***	0.06	0.77	1243 (107)	0.73***	0.06	0.73	1996 (65)	0.96	0.09	0.96
Comparable spells only	8021 (136)	0.82***	0.05	0.82	2642 (92)	0.75***	0.05	0.75	5379 (62)	1.02	0.07	1.02

(Continues)

TABLE 3 (CONTINUED)

Group	All				Consumption				Income			
	Obs.	Coeff	S.E.	r^2	Obs.	Coeff	S.E.	Error	Obs.	Coeff	S.E.	Error
Comparable, bal, 1-3 year spells	2569 (87)	0.82***	0.06	0.82	808 (44)	0.72***	0.06	0.72	1761 (55)	1.02	0.08	1.02
Comparable, bal, 4-6 year spells	1969 (88)	0.79***	0.06	0.79	613 (46)	0.70***	0.06	0.70	1356 (56)	1.03	0.07	1.03
1980s	104 (26)	0.99	0.20	0.99	29 (8)	0.92	0.20	0.92	75 (19)	1.00	0.18	1.00
1990s	493 (84)	0.55***	0.09	0.55	187 (60)	0.51***	0.09	0.51	306 (42)	0.63***	0.12	0.63
2000s	2097 (118)	0.88*	0.06	0.88	851 (77)	0.82***	0.06	0.82	1246 (54)	1.17*	0.10	1.17
2010s	2481 (103)	0.65***	0.09	0.65	741 (76)	0.51***	0.09	0.51	1740 (58)	0.98	0.11	0.98
Low income	474 (49)	0.78	0.17	0.78	460 (48)	0.79	0.17	0.79	14 (3)	0.70	0.26	0.70
Lower Middle Income	2676 (77)	0.64***	0.10	0.64	1363 (65)	0.60***	0.10	0.60	1313 (18)	0.81	0.25	0.81
Upper Middle Income	3135 (50)	0.99	0.08	0.99	999 (35)	0.90	0.08	0.90	2136 (28)	1.28*	0.15	1.28
High Income	5955 (41)	0.94	0.09	0.94	817 (30)	0.79**	0.09	0.79	5138 (39)	0.97	0.09	0.97
East Asia & Pacific	1129 (19)	0.65***	0.12	0.65	952 (15)	0.66***	0.12	0.66	177 (6)	0.77***	0.09	0.77
Europe & Central Asia	8501 (48)	0.86*	0.08	0.86	3301 (46)	0.81**	0.08	0.81	5200 (33)	1.08	0.08	1.08
Latin America & Caribbean	4422 (23)	0.92	0.21	0.92	133 (4)	0.12***	0.21	0.12	4289 (22)	0.99	0.21	0.99
Middle East & North Africa	338 (12)	0.49***	0.10	0.49	212 (10)	0.45***	0.10	0.45	126 (3)	0.55	0.36	0.55
North America	1114 (2)	0.63*	0.19	0.63	496 (1)	0.31***	0.19	0.31	618 (2)	0.72**	0.11	0.72
South Asia	175 (7)	0.83	0.15	0.83	175 (7)	0.83	0.15	0.83				
Sub-Saharan Africa	410 (44)	0.75*	0.13	0.75	409 (44)	0.77*	0.13	0.77				

Notes: The panels shows results for no-constant regressions of growth rates from surveys on growth rates from national accounts, as done by Ravallion (2003). Robust standard errors clustered at the country level (based on robust sandwich estimator). Number of countries (clusters) are shown in parentheses next to the number of observations. Statistical significance of difference from *one*, denoted by $*P \leq 0.1$, $**P \leq 0.05$, $***P \leq 0.01$.

when developing global poverty scenarios for 2030 (see Jolliffe *et al.*, 2014 for details). The results indicate a stronger correlation between growth in surveys and national accounts for longer and more comparable growth spells. For consumption means in surveys, we observe that most coefficients are significantly less than one, suggesting a lower pass through of growth from national accounts to surveys, than for income means in surveys where the coefficients are larger and the difference from unity mostly statistically insignificant. The Middle East and North Africa region stands out as a region with particularly poor correlation between national accounts and survey growth rates. We find a much stronger relationship in Eastern Europe and Central Asia than Ravallion's (2003) assessment, which mainly used data from the transition period of the early 1990s, when both national accounts and survey data were particularly poor and periods of high inflation gave large measurement errors.

The sample of comparable growth spells available in the Global Database of Shared Prosperity, which is designed to assess growth in comparable household surveys over spells of 3 to 7 years, shows clearly how large the difference in annualized growth rates can be even for modest spells length with high quality and comparable survey data. For the most recent version of the database covering spells from approximately 2013 to 2018, the gaps between annualized real growth rate in survey mean and HFCE mean range from −8.9 to 6.7 percentage points, with mean gap of −0.27 percentage points, but not statistically significantly different from zero at 95%-confidence level. When using the no-constant regression approach to assess the gaps with the Shared Prosperity Database, the regression coefficient is 0.86 for GDP (not statistically significantly different from 1) and 0.81 for HFCE (statistically significantly different from 1).

4. Implications for Global Measures of Poverty and Inequality

With the existing literature on why the gaps exist and the empirical evidence on the size of the gaps in mind, we consider two approaches for adjusting survey data to align more closely with national accounts estimates. We examine the implications of each approach for global poverty and inequality measures. One simulation uniformly scales up the welfare vector (i.e. the consumption or income vector) from survey data to match survey means with national accounts estimates. While this scaling approach has been frequently used, notably by Pinkovskiy and Sala-i-Martin (2014, 2016), we believe that the distribution-neutral adjustment is based on an untenable assumption and unsupported by the empirical evidence on measurement error in surveys. We do nonetheless examine these simulations in part as a point of comparison with existing literature and, also as a point of contrast with our main simulation.

For the second simulation, we assume that the HHS–NAS gap is mainly a result of the incomes and consumption of better-off households being inadequately captured in survey data (primarily in terms of item nonresponse and underreporting). This simulation draws on the empirical evidence discussed in Section 3, that measurement error in surveys disproportionately comes from underestimating the top tail of the distribution. Our approach is informed by

the work of Lakner and Milanovic (2016) and closely follows Chandy and Seidel (2017a, 2017b) who adjust the top tail of each survey distribution, in proportion to the HHS–NAS gap. Both fit a Pareto distribution to allocate the HHS–NAS gap to the richest household of the survey distribution. Their methods build on an approach suggested by Atkinson (2007) who uses a Pareto imputation to "elongate" the upper part of the distribution. Lakner and Milanovic add the full HHS–NAS gap to the top decile of the distribution in their data (though with an upper bound). We follow the approach of Chandy and Seidel (2017b), who add a top segment to the Lorenz curve from the survey distribution, with income corresponding to *half* the HHS–NAS gap, fitted with the pareto distribution from the top survey decile. In this method, the elongation of the distribution is a function of both the size of the gap and the observed inequality of the top survey decile.

4.1. *Poverty*

Several researchers have proposed uniformly scaling up survey data to match national accounts. Bhalla (2002) scales up survey means to match HFCE from national accounts to estimate global poverty and inequality. Bourguignon and Morrisson (2002), Sala-i-Martin (2006), Sala-i-Martin and Pinkovskiy take a similar approach but scale up to match per capita GDP. Chen and Ravallion (2010) offer a more measured approach to scaling up survey distributions by allowing the scaling factor to be informed by both national accounts and survey means.[16]

Since per capita income and consumption in national accounts are generally higher than in surveys, scaling up survey data in proportion to the gap leads to lower estimated poverty rates, when holding the international poverty line and the distributions the same. This is essentially true by assumption. Compared to our survey estimates, on average, country-level poverty estimates at the international poverty line (IPL) of $1.90 are on average 34 percent lower using HFCE, and 61 percent lower using GDP. We analyze the differences for household surveys where extreme poverty is greater than 3 percent. At poverty rates lower than this, small changes can lead to very large changes in percentage terms. Figure 5, Panel A, compares survey estimates of poverty at the $1.90 line to corresponding measures using national accounts means combined with distributions from surveys. Only in 11 percent of observations is poverty higher using HFCE rather than using survey estimates, and only 2.5 percent in the case of GDP. This illustrates the much lower level of poverty resulting from using national accounts means in combination with survey distributions and the $1.90 line. Naturally, when aggregated to global estimates, poverty measures using national accounts means is much lower compared to that using survey means, as seen in the right pane of Panel A of Figure 5. Our estimates using national accounts means and survey distributions suggest that the World Bank's

[16]More specifically, they scale up survey means to correspond with the average of the observed survey mean and a predicted survey mean based on a regression with national accounts estimates as the explanatory variable.

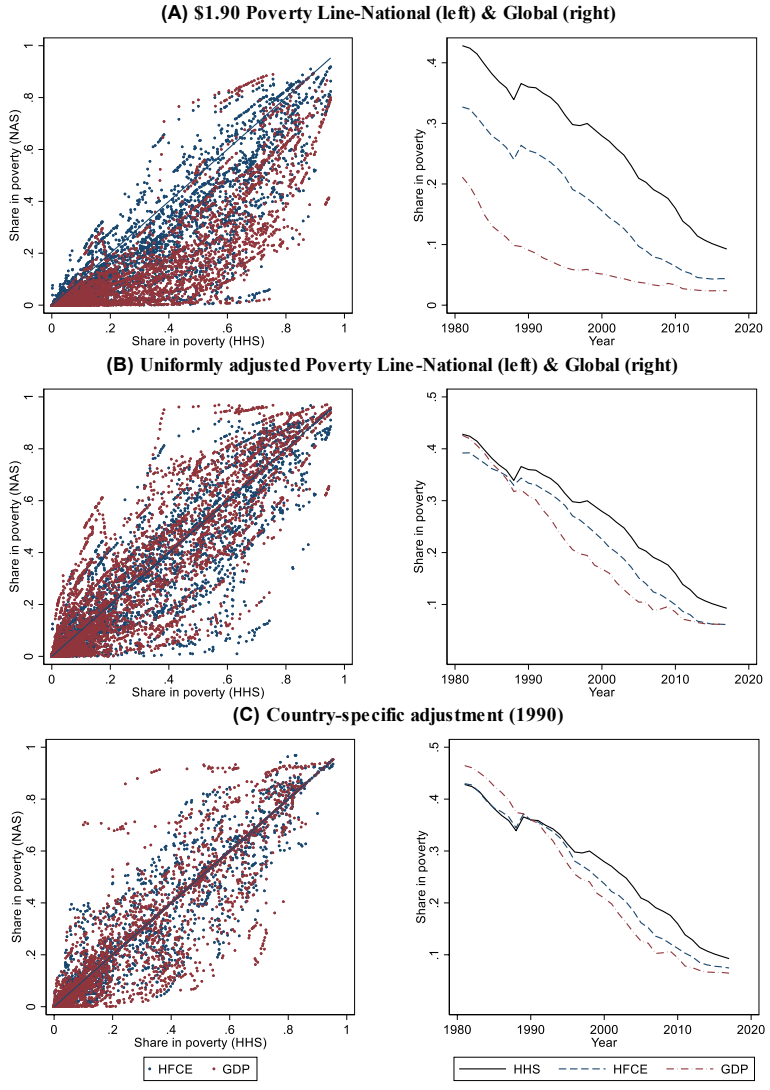


Figure 5. Poverty Measures from National Accounts vs Household Surveys.

Notes: Panels compare the poverty estimates using various welfare measures and international poverty lines using the World Bank's PovcalNet database with annual national and global poverty estimates from 1991 to 2017. Panel A compares poverty at the $1.90/day international poverty line (IPL) using survey means, HFCE means and GDP means. The left charts in all three panels shows unadjusted survey estimates on vertical axis versus national accounts-based measures on horizontal axis. Observations on the 45-degree indicate estimates from surveys and national accounts are similar and observations below (above) indicate that estimates from national accounts are lower (higher) than those from surveys. The right-hand chart show global aggregate for the three measures of poverty over time. Panel B shows the same welfare measures, but with global poverty lines used for HFCE and GDP adjusted by the average gap between survey and national accounts means. Panel C uses country specific poverty lines adjusted by the country-specific HHS–NAS gap in 1990. [Colour figure can be viewed at wileyonlinelibrary.com]

3 percent global poverty target for 2030 was reached by year 2011 in the case of GDP, and is estimated at 5.7 percent when using HFCE the same year. These findings are broadly consistent with Sala-i-Martin and Pinkovskiy (2016) who calculated measures up to 2010 using 2005 PPPs and the $1.25-line. It is worth noting that with NAS measures being much higher than survey measures at any point in time, any starting point for poverty trajectories based on national accounts are also much lower than for survey estimates. For example, estimates for 1990, the baseline year of the UN Millennium Development Goals, show global poverty measured using GDP at 9 percent, 25 percent with HFCE, and above 35 percent using survey means.

The $1.90 poverty line may not be the appropriate threshold for international poverty measurement when the distribution is adjusted to the level of national accounts means. Since the $1.90 line is itself estimated from national poverty lines originating from household surveys (see Ferreira *et al*., 2016; Jolliffe and Prydz, 2016), the poverty line is likely also underestimated, and therefore not necessarily suitable for use with living standards measured using national accounts. A simple adjustment of the international poverty line for use with national accounts, would be to scale the poverty line up by the same proportion as the survey mean (to equate with national accounts).[17] This would give a poverty line of $2.38 for use with HFCE and $3.55 for use with GDP. These estimates are based on the overall ratio observed in Table 1. For HFCE, the gap is −0.202, so adjusting the poverty line of 1.90, would give 1/(1−0.202)*1.9 = 2.38. For GDP, the gap is 0.465, resulting in adjusted poverty line of 1/(1−0.465)*1.9 = 3.55. Figure 5, Panel B illustrates the poverty estimate for each country using these poverty lines. Although poverty estimates with the adjusted lines are more closely aligned to survey estimates than in Panel A, there is still substantial variation, resulting from the variation in the gaps across countries combined with the uniformly adjusted poverty lines.

Just as it does not make much sense to scale up survey means and not change the $1.90 poverty line, it is also not reasonable to increase the poverty line by the same proportion as the survey means. Part of the poverty line is based on estimating the cost of obtaining minimum nutrition needs and this estimate need not necessarily be affected by underreporting of consumption (or income). Chen and Ravallion (2010) clarify this point better by noting that typical methods for setting national poverty lines will underestimate the poverty line if non-food spending is underestimated in surveys, and thus any correction for the underestimation of non-food spending would also lead to higher poverty lines. But this adjustment, by construction, would almost certainly be less than the entire gap between survey means and national accounts.

A different way of adjusting the IPL for poverty measurement based on national accounts means, is to set a country-specific IPL is based on the country-specific

[17]Since a proportion of the national poverty lines for many poor countries is based on the pricing a basket of caloric assumption it can be argued that this proportion is not underestimated to the same extent as non-food consumption or income. Thus, simply adjusting the poverty line by the average gap may be too drastic, as it would imply that both food and non-food is underestimated in surveys used to define the poverty line. Moreover, we are using the average the gap, which is larger than the gaps typically found in low-income countries.

survey-national accounts gap observed in the data. Alternatively (and equivalently), one could use the national accounts series, scaled down by the gap, and use the $1.90 line. Such an approach could be justified from a perspective that national accounts may be a more comparable measure of changes in living standards over time, but that surveys, on average, do a better job at measuring both inequality and the level of poverty. Under this approach, large changes to survey methodology that affect the survey mean (but smaller changes to inequality), would have a much smaller effect on the evolution of poverty which would be based on a more stable series of national accounts means. Such an adjustment can be done for a particular year, or by taking the average ratio for a country over a longer time period. We estimate such country specific IPLs for use with national accounts for 1990, ensuring that poverty estimates from national accounts and surveys are aligned in that year. The results for global poverty are shown in Panel C of Figure 5. While the poverty estimates are equal (by design) in 1990, there is considerable variation over time, and, in global measures the rate of decline is much larger, due to the faster growth rates of national accounts means compared with survey means.

Across all the methods which involve substituting survey means with national accounts means, poverty is estimated to be lower and falling faster when compared to traditional survey measures. The more rapid decline seen in the poverty measures using national accounts points to a concern with current use of national accounts growth rates in extrapolating household survey estimates for years with missing surveys. Even if household survey means are used for poverty estimation for survey years, national accounts growth rates are used to interpolate such estimates to non-survey years and for nowcasts and projections of poverty in the future. Current methods use actual or projected national accounts growth rates to align poverty estimates to non-survey years for global aggregation. Removing the bias implicit in this method, suggests a slower global decline in poverty than the World Banks official poverty numbers.

Because India, Indonesia and China, countries which historically have been the home of a large share of the global poor, historically have had household surveys for most reference years for which the World Bank reports poverty, the effect on global numbers is of less concern. However, the lack of recent surveys available for India, the home to a large share of the world's poor, has generated greater uncertainty about poverty estimates from national accounts based extrapolation of the latest available survey.[18] Extrapolations or projections of poverty beyond the World Bank's latest reference year that use national accounts growth rates, commonly use an adjustment factor to adjust for the discrepancy in growth rates between national accounts and surveys. But even if the systematic bias and overall error is reduced by applying the adjustment factors, the precision of the method is still poor as reflected in the relatively high measures of error in Table 2, suggesting

[18]In recent years though, no survey has been available for India, making global poverty estimates for these years highly uncertain. Nowcasting approaches that take into account the limited pass through of national accounts growth to household consumption growth, have been tested. For a closer assessment of the current method applied by the World Bank for the situation in India see Newhouse and Vyas (2019).

that national accounts growth is a poor predictor of survey growth and thus adding uncertainty about poverty projections.

The discussion above assumes a distribution-neutral adjustment to account for the HHS–NAS gap. However, "top income" adjustments are more appropriate if the source of the gap is originating from top incomes being mismeasured in surveys due to biased response rates or underreporting by the richest households. Since this adjustment mainly effects the very top of the distribution, that in all countries is above the $1.90 threshold, it has little effect on poverty measures. However, since we add a population segment to the survey distribution, poverty measures fall proportionally to the number of observations added to the top of the distribution. For example, in a country where poverty is estimated to be 30 percent in an unadjusted survey, and we expand the distribution by adding 10 percent of the population to adjust for missing people the top, the poverty rate would fall to 27.2 percent. Figure A2.1 (in Appendix 2) shows the global poverty trends for the adjusted and unadjusted HHS data, with very similar estimates. The adjusted distributions give slightly lower estimates due to the added number of people in the denominator in poverty estimates.

4.2. *Inequality and Inclusive Growth*

While the implications of the top income adjustment have little effect on poverty measures, the implications for levels and trends in inequality are significant. Figure 6 compares the Gini coefficients from the unadjusted distributions with inequality from the distributions with top-adjustments, based on adding a top segment to the Lorenz curve from the survey distribution as proposed by Chandy and Siedel (2017b). (We thank the authors for providing the replication code to implement this.) At the national level (shown in Panel A of Figure 6), the observed Gini coefficients are on average 20 percent higher when using the top-income adjusted distributions. Other measures of levels of inequality, such as the 90/10 ratio and the Palma ratio would also be drastically affected by such adjustments.

Measuring inequality globally using the top-adjusted distributions also results in a much higher level of global inequality, with the top income adjustment increasing the global Gini in 2017 from 62 to 67.[19] Panel B of Figure 6 shows trends that are broadly similar for the adjusted and unadjusted surveys, showing a robust decline in global inequality since about year 2000. The levels and patterns are in line with those of Lakner and Milanovic (2016) who use a different form of adjustment and aggregation, but only provide estimates until 2008. For 2008 they estimate a global Gini of 67.0 using 2011 PPPs, while our estimates are 66.7.[20] Their top-income adjusted estimates are 2.9 to 6.3 percentage points higher for 2008, within the same range as our top-adjusted estimate,

[19]We measure global interpersonal inequality, capturing inequality of individual incomes (or consumption), referred to as "concept 3" inequality by Milanovic (2005). We estimate a global Gini coefficient giving each individual equal weight, regardless of where they live, using the same weights as used for estimating global poverty. The global distribution is generated using the reference year distributions for global poverty measurement in PovcalNet, using the method described in Section 2.

[20]Lakner and Milanovic (2016) conduct most of their analysis using 2005 PPPs, but provide estimates based on 2011 PPPs as well.

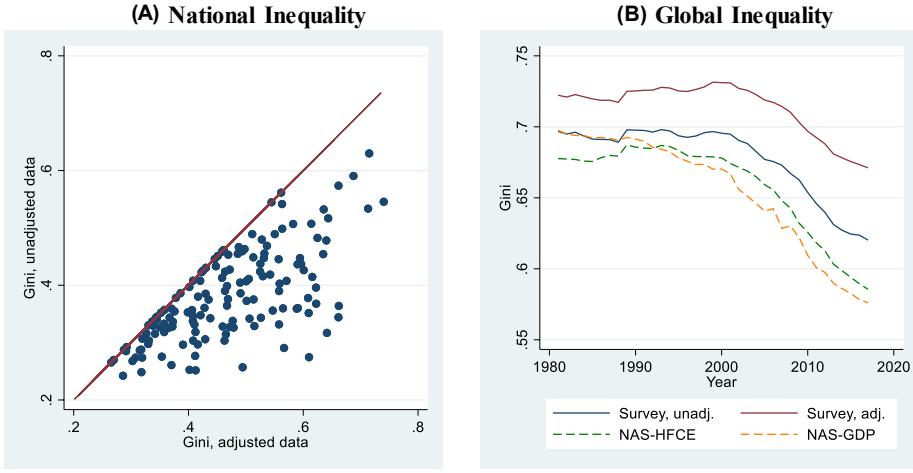


Figure 6. Inequality (Gini) Unadjusted and Adjusted Measures.
Notes: Panel A compares Gini coefficients estimated at the national level for the unadjusted survey data (on vertical axis) and from top-income-adjusted data on horizontal axis for the most recent survey for each country. Estimates on the 45-degree line indicate identical estimates for the adjusted and unadjusted data. Panel B shows the global Gini calculated using four different distributions. The solid blue line shows the global Gini using unadjusted survey data, while the solid maroon line shows the global Gini using the top income adjusted survey data. The dashed lines show global Gini using HFCE and GDP means from national accounts in combination with unadjusted distributions from surveys. [Colour figure can be viewed at wileyonlinelibrary.com]

which is 4.3 percentage points higher for 2008. Using the unadjusted distributions with national accounts means instead of survey means, global inequality is lower and declining faster. This is expected from the pattern of national accounts measures, which are larger and growing faster than survey measures, especially in middle-income countries. Notably, global inequality has continued falling rapidly using all measures. Our survey-based estimates have fallen 6 to 7 percentage points since 2000.

Generally, the larger HHS–NAS gap among middle-income countries leads to relatively larger adjustments to inequality measures in these countries. These systematic differences along the income gradient of countries also lead to insights into the relationship between economic development and levels of inequality. Evidence of the cross-sectional Kuznets curve—the hypothesis that income inequality first increases and then declines with development—has recently been questioned. Palma (2011, p. 87) suggests that the "the 'upwards' side of the 'Inverted-U' between inequality and income per capita has evaporated." This is indeed true for the unadjusted Gini coefficients in our sample. However, Gini coefficients from the top-income adjusted data, suggest that there is an upwards sloping segment and firmly re-establish a Kuznets-like relationship between economic development and observed inequality, as seen in Figure 7. The clearly inverted U-shaped curve results from the lowess regression of the adjusted Ginis. A quadric form regression of the Gini estimates from the

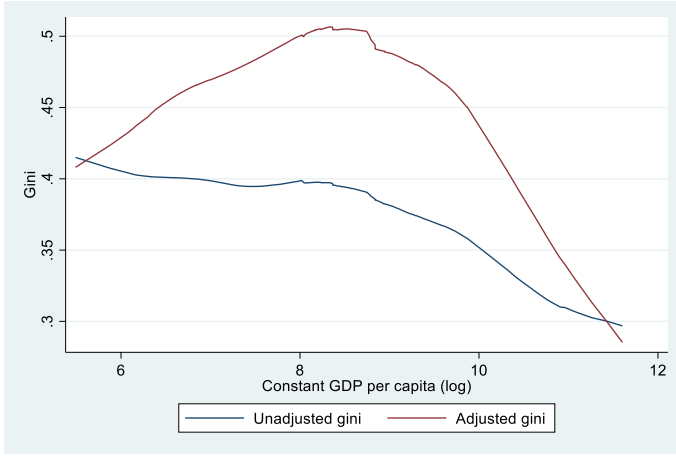


Figure 7. The "Kuznets Curve": Inequality and Economic Development
Notes: Lines show results from lowess non-parametric regressions (with bandwidth of 0.8) of the unadjusted and top-income adjusted ginis on log GDP per capita. The last observation for each country is used. To test for the presence of an inverse u-relationship, we utilize the Stata command UTEST (Lind and Mehlum, 2010), which tests the hypothesis that the relationship is increasing at the start of the interval and decreasing at the end. The test confirms an inverse u-shape that is strongly statistically significant ($P < 0.01$) for the relationship between the adjusted ginis and log GDP per capita, and statistically insignificant ($P > 0.1$) for the unadjusted ginis. We estimate a quadratic specification of the relationship between the observed ginis and log GDP per capita. [Colour figure can be viewed at wileyonlinelibrary.com]

adjustment on GDP per capita explains more than twice as much variation as the regression of the unadjusted measures.

Measures of the degree to which economic growth is inclusive are also affected by the assumption that at least some of the HHS-NAS gap originates from missing top incomes in survey data. In measuring the degree to which growth is inclusive, the World Bank's twin goals and the UN Sustainable Development Goals (SDGs) monitor growth in income and consumption of the bottom 40 percent in each country, relative to the growth for the overall population (World Bank, 2016). For a recent set of comparable spells used to monitor these goals, 56 out of 88 countries (64 percent) that are available in the Shared Prosperity Database for 2013 to 2018, reported a positive "shared prosperity premium": the growth of mean income or consumption among the bottom 40 percent exceeded that of the overall mean growth. Panel A of Figure 8 shows the growth of the bottom 40 percent on the vertical access, and of the overall mean on the horizontal axis. The majority of observations are above the 45-degree line, indicating that growth was inclusive, and inequality, by this measure, was falling. However, these estimates rely solely on survey data. If we assume that the gap between surveys and national accounts is partly due to missing top incomes in surveys, it may be justifiable to compare the growth of the bottom 40 percent from surveys with the mean of national accounts, assuming that this better reflects overall growth. Under such a comparison, only 40 countries (49 percent of the 81 countries for which we have HFCE data for the

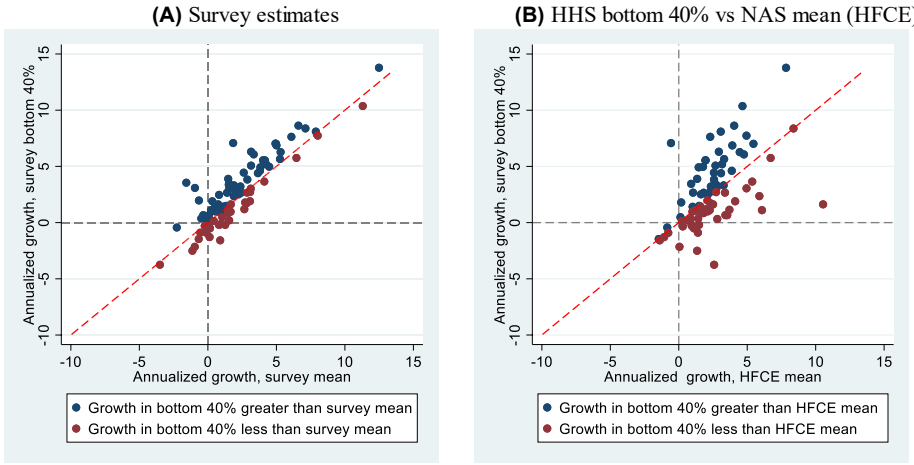


Figure 8. Shared Prosperity (Growth in Bottom 40 percent vs Growth in Mean).
Notes: Panel A shows the growth from surveys in the bottom 40 percent on the vertical access, and growth of the in the mean on the horizontal axis, using the latest data from the 2013–2018 Shared Prosperity Database. The 45-degree line indicates equal growth between the bottom 40 percent and the mean. Of the 81 observations available in both the Shared Prosperity and NAS database, 52 countries (64 percent) have higher growth in the bottom 40 percent than in the overall mean, indicating reduction in inequality by this measure. Panel B compares growth in the bottom 40 percent from surveys on the vertical axis with growth in mean consumption as measured by national accounts. In this case only 40 countries (49 percent) show the bottom 40 percent growing faster than the mean as measured by HFCE. [Colour figure can be viewed at wileyonlinelibrary.com]

spells) had a positive shared prosperity premium where the bottom 40 percent grew faster than the mean as measured by HFCE. Panel B of Figure 8 illustrates the relationship and differences between growth in the bottom 40 percent and growth in the mean with the adjusted distributions. The population-weighted average annualized shared prosperity premium for the period falls by 0.6 percentage points, from a positive 0.5 percent growth to −0.1 percent, inverting the positive global picture of trends in inequality measured by the official spells. This further illustrates that conventional measures of development can be misleading if one assumes the HHS–NAS gap to originate from a lack of capture of consumption and income of the richest households in surveys.

5. Conclusion

This paper has compiled a large new data set for assessing the correspondence between per capita monetary living standards measured in national accounts and household surveys. The data show that the gaps in measurement across the two data sources are larger than in previous assessments. Our assessment concludes that the gap does not seem to be due to survey income versus consumption measures, as suggested by Deaton (2005) and Ravallion (2003). Rather, the average gaps are closely aligned with the level of economic development, with gaps being largest for middle-income countries, both in terms of levels and growth rates. The gaps, and corresponding implications for poverty and

inequality measures, are largest for middle-income countries which have experienced periods of rapid growth. With the large majority of the extreme global poor currently living in middle-income countries, the implications of these gaps for measuring and understanding the evolution of global poverty and inequality aggregates are large.

It is increasingly documented that the HHS–NAS gap at least partly originates from the inability of surveys to capture the full level of consumption and income for all households. This paper illustrates the potential implications for common poverty and inequality measures from adjusting survey data based on differing assumptions of the source of the gap. Under a distribution-neutral gap scenario, which would justify substituting survey means with national accounts means, global poverty would be much lower using the $1.90 line, and the Sustainable Development Goal of "ending" extreme poverty would already be close to being met, or at least be easily within reach. This is in line with findings of Sala-i-Martin and Pinkovskiy (2016) that uses this method to measure global poverty. However, we argue that when using national accounts for measuring poverty, the $1.90 line should be adjusted. Using an international poverty line adjusted for systematic differences between surveys and national accounts would result in global poverty measures that are more in line with survey-only measures, but still cause relatively large changes to country level measures. Regardless of the poverty line used with national accounts, the rate of poverty reduction is greater than that measured in surveys because national accounts growth rates are typically higher. This last observation illustrates the perils of using growth rates from national accounts to extrapolate global poverty, which is the current practice. The current approach likely exaggerates the decline in poverty estimates when no survey data is available.

The scenario which assumes that the HHS-NAS gap is due to surveys not fully capturing consumption or incomes of the richest households in societies, and therefore makes adjustments to the top segment of survey distributions, has small implications for poverty measures, but drastic implications for typical inequality measures. Adjusting survey data for missing top incomes to account for part of the NAS-HHS gap increases national and global inequality considerably, as measured by the Gini coefficient. Moreover, the hypothesis that as economies develop, inequality first increases and then decreases, also known as the Kuznet's curve, is much more strongly supported in our cross-sectional sample of "top income"-adjusted distributions.

Because of the large gaps between survey and national accounts data, and particularly the large variation in gaps across countries and over time, the prospect of reliably filling gaps in poverty data with estimates of poverty imputed from national accounts growth rate or aggregates, is limited. The errors (differences) of estimates based on national accounts data, in comparison to survey data, are very large. As long as household surveys appear to be the preferred method of measuring poverty and inequality, national accounts data offer only partial hope for filling data gaps. Ultimately, more frequent and properly sampled household surveys—designed to capture the full incomes and consumption of all households, potentially with integration of tax records and administrative data as countries get richer—appears to be the best approach for improving our understanding of poverty and inequality.

References

Abdelkrim, A. and J.-Y. Duclos, *DASP: Distributive Analysis Stata Package*, PEP, World Bank, UNDP and Université Laval, Washington, DC, 2007.

Anand, S. and P. Segal, "What Do We Know About Global Income Inequality?" *Journal of Economic Literature*, 46, 57–94, 2008.

Angrist, N., P. Goldberg, and D. Jolliffe, "Why is Growth in Developing Countries so Hard to Measure?" *Journal of Economic Perspectives*, 35, 215–42, 2021.

Artis, M. and M. Marcellino, "Fiscal Forecasting: The Track Record of the IMF, OECD and EC," *Econometrics Journal*, 4, 20–36, 2001.

Aruoba, S. B., F. X. Diebold, J. Nalewaik, F. Schorfheide, and D. Song, "Improving GDP Measurement: A Measurement-error Perspective," *Journal of Econometrics*, 191, 384–97, 2016.

Atamanov, A., R. A. Castaneda Aguilar, C. Diaz-Bonilla, D. Jolliffe, C. Lakner, D. G. Mahler, J. Montes, M. Herrera, L. Liliana, D. Newhouse, M. C. Nguyen, E. B. Prydz, P. Sangraula, S. A. Tandon, and J. Yang, "September 2019 PovcalNet Update: What's New," *Global Poverty Monitoring Technical Note*, 10, World Bank, Washington, DC, 2019.

Atkinson, A. B., "Measuring Top Incomes: Methodological Issues," in A. B. Atkinson and T. Piketty (eds), *Top Incomes over the Twentieth Century: A Contrast between Continental European and English-Speaking Countries*, Oxford University Press, Oxford and New York, 18–42, 2007.

Badiani, R., B. Baulch, L. Brandt, V. H. Dat, N. T. Giang, J. Gibson, J. Giles, I. Hinsdale, P. Hung, and V. Kozel, *2012 Vietnam Poverty Assessment: Well Begun, Not Yet Done-Vietnam's Remarkable Progress on Poverty Reduction and the Emerging Challenges*, World Bank, Washington, DC, 2013. https://openknowledge.worldbank.org/handle/10986/12326

Banerjee, A. and T. Piketty, "Top Indian Incomes, 1922–2000," *World Bank Economic Review*, 19, 1–20, 2005.

Beegle, K., J. De Weerdt, J. Friedman, and J. Gibson, "Methods of Household Consumption Measurement Through Surveys: Experimental Results From Tanzania," *Journal of Development Economics*, 98, 3–18, 2012.

Bhalla, S. S., *Imagine There's No Country: Poverty, Inequality, and Growth in the Era of Globalization*, Peterson Institute for International Economics, Washington, DC, 2002. https://www.piie.com/bookstore/imagine-theres-no-country-poverty-inequality-and-growth-era-globalization

Bourguignon, F. and C. Morrisson, "Inequality Among World Citizens: 1820–1992," *American Economic Review*, 92, 727–744, 2002.

Browning, M., T. F. Crossley, and J. Winter, "The Measurement of Household Consumption Expenditures," *Annual Review of Economics*, 6, 475–501, 2014.

Buddenberg, D. and W. A. Byrd, *Afghanistan's Drug Industry: Structure, Functioning, Dynamics, and Implications for Counter-Narcotics Policy*, United Nations Office on Drugs and Crime and World Bank, New York and Washington, DC, 2006.

Chandy, L. and B. Seidel, "How Much Do We Really Know About Inequality Within Countries Around the World? Adjusting Gini coefficients for missing top incomes," Brookings Institution, 2017a. https://www.brookings.edu/opinions/how-much-do-we-really-know-about-inequality-within-countries-around-the-world/

_____, "Allocating Income from Survey-National Accounts Gap to Top Incomes According to Pareto Distribution," Brookings Institution, 2017b. https://www.brookings.edu/wp-content/uploads/2017/02/technical-appendix-02-13-17.pdf

Charmes, J., "The Informal Economy Worldwide: Trends and Characteristics," *Margin: Journal of Applied Economic Research*, 6, 103–132, 2012.

Chen, S. and M. Ravallion, "The Developing World is Poorer than We Thought, But No Less Successful in the Fight Against Poverty," *Quarterly Journal of Economics*, 125, 1577–1625, 2010.

Damgaard, J., T. Elkjaer, and N. Johannesen, "What Is Real and What Is Not in the Global FDI Network?," IMF Working Papers 19, 2019.

Deaton, A., "Measuring Poverty in a Growing World (or Measuring Growth in a Poor World)," *Review of Economics and Statistics*, 87, 1–19, 2005.

Deaton, A. and V. Kozel, "Data and Dogma: The Great Indian Poverty Debate," *World Bank Research Observer*, 20, 177–199, 2005.

Deaton, A. and S. Zaidi, "Guidelines for Constructing Consumption Aggregates for Welfare Analysis," Living Standards Measurement Study, Working Paper no. 135, World Bank, Washington, DC, 2002.

Engle-Stone, R., C. Sununtnasuk, and J. L. Fiedler, "Investigating the Significance of the Data Collection Period of Household Consumption and Expenditures Surveys for Food and Nutrition Policymaking: Analysis of the 2010 Bangladesh Household Income and Expenditure Survey," *Food Policy*, 72, 72–80, 2017.

Farfán, G., M. E. Genoni, and R. Vakis, "You Are What (and Where) You Eat: Capturing Food Away from Home in Welfare Measures," *Food Policy*, 72, 146–156, 2017.
Ferreira, F. H. G., S. Chen, A. Dabalen, Y. Dikhanov, N. Hamadeh, D. Jolliffe, A. Narayan, E. B. Prydz, A. Revenga, P. Sangraula, U. Serajuddin, and N. Yoshida, "A Global Count of the Extreme Poor in 2012: Data Issues, Methodology and Initial Results," *Journal of Economic Inequality*, 14, 141–172, 2016.
Gibson, J., K. Beegle, J. D. Weerdt, and J. Friedman, "What Does Variation in Survey Design Reveal About the Nature of Measurement Errors in Household Consumption?" *Oxford Bulletin of Economics and Statistics*, 77, 466–474, 2015.
Holden, K. and D. A. Peel, "On Testing for Unbiasedness and Efficiency of Forecasts," *The Manchester School of Economic and Social Studies*, 58, 120–127, 1990.
Jäntti, M., V. M. Törmälehto, and E. Marlier, "The Use of Registers in the Context of EU-SILC: Challenges and Opportunities," Eurostat Statistical Working Papers, Publications Office of the European Commission, 2013.
Jolliffe, D., "Measuring Absolute and Relative Poverty: The Sensitivity of Estimated Household Consumption to Survey Design," *Journal of Economic and Social Measurement*, 27, 1–23, 2001.
Jolliffe, D., P. Lanjouw, S. Chen, A. Kraay, C. Meyer, M. Negre, E. B. Prydz, R. Vakis, and R. Wethli, *A Measured Approach to Ending Poverty and Boosting Shared Prosperity: Concepts, Data, and the Twin Goals*, World Bank, Washington, DC, 2014.
Jolliffe, D. and E. B. Prydz, "Estimating International Poverty Lines from Comparable National Thresholds," *Journal of Economic Inequality*, 14, 185–198, 2016.
Karshenas, M., "Measurement and Nature of Absolute Poverty in Least Developed Countries," Economic Working Paper Series no. 129, School of Oriental and African Studies, University of London, London, 2001.
_____, "Global Poverty: National Accounts Based versus Survey Based Estimates," *Development and Change*, 34, 683–712, 2003.
Korinek, A., J. A. Mistiaen, and M. Ravallion, "Survey Nonresponse and the Distribution of Income," *Journal of Economic Inequality*, 4, 33–55, 2006.
Kouame, C. Y., B. B. Kalimili, and F. Pirlea, "Many African Economies are Larger than Previously Estimated," *World Development Indicators*, 2019. http://datatopics.worldbank.org/world-development-indicators/stories/many-economies-in-ssa-larger-than-previously-thought.html
Kuznets, S., "Economic Growth and Income Inequality," *American Economic Review*, 45, 1–28, 1955.
Lakner, C., D. G. Mahler, M. Negre, and E. B. Prydz, "How Much Does Reducing Inequality Matter for Global Poverty?" *Policy Research Working Paper* No. 8869, World Bank, Washington, DC, 2019. https://openknowledge.worldbank.org/handle/10986/31796
Lakner, C. and B. Milanovic, "Global Income Distribution: From the Fall of the Berlin Wall to the Great Recession," *World Bank Economic Review*, 30, 203–232, 2016.
Lind, J. T. and H. Mehlum, "With or Without U? The Appropriate Test for a U-shaped Relationship," *Oxford Bulletin of Economics and Statistics*, 72, 109–118, 2010.
Milanovic, B., *Worlds Apart: Measuring International and Global Inequality*, Princeton University Press, Princeton, NJ, 2005.
Newhouse, D. L. and P. Vyas, "Estimating Poverty in India without Expenditure Data: A Survey-to-Survey Imputation Approach," *Policy Research Working Paper* No. 8878, World Bank, Washington, DC, 2019. https://openknowledge.worldbank.org/handle/10986/31868
Nolan, B., M. Roser, and S. Thewissen, "GDP Per Capita Versus Median Household Income: What Gives Rise to the Divergence Over Time and How Does This Vary Across OECD Countries?" *Review of Income and Wealth*, 65, 465–494, 2019.
OECD, *Measuring the Non-Observed Economy: A Handbook*, OECD Publishing, Paris, 2002.
Palma, J. G., "Homogeneous Middles vs. Heterogeneous Tails, and the End of the 'Inverted-U': the Share of the Rich is What it's All About," *Cambridge Working Papers in Economics*, no. 1111, Faculty of Economics, University of Cambridge, 2011.
Pinkovskiy, M. and X. Sala-i-Martin, "Africa Is On Time," *Journal of Economic Growth*, 19, 311–338, 2014.
_____, "Lights, Camera… Income! Illuminating the National Accounts-Household Surveys Debate," *Quarterly Journal of Economics*, 131, 579–631, 2016.
Pradhan, M., "Welfare Analysis with a Proxy Consumption Measure: Evidence from a Repeated Experiment in Indonesia," *Fiscal Studies*, 30, 391–417, 2009.
Ravallion, M., "Measuring Aggregate Welfare in Developing Countries: How Well Do National Accounts and Surveys Agree?" *Review of Economics and Statistics*, 85, 645–652, 2003.
Sala-i-Martin, X., "The World Distribution of Income: Falling Poverty and… Convergence, Period," *Quarterly Journal of Economics*, 121, 351–397, 2006.
Schündeln, M., "Multiple Visits and Data Quality in Household Surveys," *Oxford Bulletin of Economics and Statistics*, 80, 380–405, 2018.

Serajuddin, U., H. Uematsu, C. Wieser, N. Yoshida, and A. Dabalen, "Data Deprivation: Another Deprivation to End," *Policy Research Working Paper* No. 7252, World Bank, Washington, DC, 2015. https://openknowledge.worldbank.org/handle/10986/21867

Stiglitz, J., A. Sen, and J. Fitoussi, "Report of the Commission on the Measurement of Economic Performance and Social Progress (CMEPSP)," 2009. https://ec.europa.eu/eurostat/documents/8131721/8131772/Stiglitz-Sen-Fitoussi-Commission-report.pdf

Stone, R., D. G. Champernowne, and J. E. Meade, "The Precision of National Income Estimates," *Review of Economic Studies*, 9, 111–125, 1942.

Subramanian, S. and D. Jayaraj, "Growth and Inequality in the Distribution of India's Consumption Expenditure: 1983 to 2009–10," *Economic and Political Weekly*, 50, 7–8, 2015.

Sundaram, K. and S. D. Tendulkar, "Poverty Has Declined in the 1990s: A Resolution of Comparability Problems in NSS Consumer Expenditure Data," *Economic and Political Weekly*, 38, 327–337, 2003.

Winter, J., "Response Bias in Survey−Based Measures of Household Consumption," *Economics Bulletin*, 3, 1–12, 2003.

World Bank, *Poverty and Shared Prosperity 2016: Taking on Inequality*, World Bank, Washington, DC, 2016. https://openknowledge.worldbank.org/handle/10986/25078

Yang, J. and M. C. Nguyen, "March 2021 Update to the Global Database of Shared Prosperity: What's New," *Global Poverty Monitoring Technical Note*, 16, World Bank, Washington, DC, 2021. https://openknowledge.worldbank.org/handle/10986/35389

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Appendix 1: Comparison with Deaton (2005) and Ravallion (2003)
Table A1.1: Comparison with Other Studies: Survey-HFCE Gap
Appendix 2: Supplemental Tables and Figure
Table A2.1: Gaps in Survey Mean and National Accounts Means, Alternative Weighting
Table A2.2: Gaps in Growth Rates, Alternative Weighting
Figure A2.1: Global Poverty for Unadjusted and Top Income Adjusted HHS