

Why Is Growth in Developing Countries So Hard to Measure?

Noam Angrist, Pinelopi Koujianou Goldberg, and Dean Jolliffe

GDP and growth estimates in developing countries are often perceived to be fraught with problems and potentially manipulated. For example, when Nigeria's government changed the way it calculated GDP in 2014, the country's official GDP grew 89 percent overnight, raising concerns that the statistics were being manipulated for political purposes (as reported in *The Economist* 2014). This GDP revision updated the base year from 1990 to 2010, reflecting the economy's changing structure and giving greater weight to growing industries like mobile technology. In China, GDP growth estimates have been routinely called into question since the mid-2000s, and recent studies estimate that official Chinese statistics overstated average annual growth by 1.8 percentage points between 2010 and 2016 (Chen et al. 2019). In India, Subramanian (2019) estimates that changes

■ Noam Angrist is a Fellow, Oxford University, Oxford, United Kingdom, Consultant, World Bank, Washington, DC, and Co-Founder, Young Love, Gaborone, Botswana. Pinelopi Koujianou Goldberg is Elihu Professor of Economics, Yale University, New Haven, Connecticut. She is also a Distinguished Fellow, Centre for Economic Policy Research, London, United Kingdom, Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts, Executive Committee Member of the Bureau for Research and Economic Analysis of Development, and Non-resident Fellow at the Peterson Institute of International Economics. Dean Jolliffe is a Senior Advisor in the Development Data Group, World Bank, Washington, DC. He is also a Research Fellow, Institute for the Study of Labor (IZA), a Fellow, Global Labor Organization, and adjunct faculty with the School of Advanced International Studies. Their email addresses are noam.angrist@bsg.ox.ac.uk, penny.goldberg@yale.edu, and djolliffe@worldbank.org.

For supplementary materials such as appendices, datasets, and author disclosure statements, see the article page at <https://doi.org/10.1257/jep.35.3.215>.

in data sources and methodology in 2011 led to an overestimation of annual growth by 2.5 percent between 2011 and 2012 and 2016 and 2017. Concerns are not confined to growth overestimation. Kerner, Jerven, and Beatty (2017) find that some lower-income countries underreport economic growth to maintain foreign assistance.

Is there robust evidence that growth is systematically mismeasured or measured less reliably in developing countries, or do such concerns reflect overgeneralizations based on a small number of widely publicized but nonrepresentative examples? More broadly, what specific challenges do developing countries face in measuring the growth of their economies? This article seeks to address these questions while offering thoughts on how growth measurement in developing countries can be improved.

There are plausible reasons why growth measurement could be more challenging for developing countries. Developing countries have lower statistical capacity, are often associated with weaker institutions and governance, have large informal sectors that are inherently hard to measure, and tend to be more reliant on agriculture. Volatile growth is harder to measure, and growth is more volatile in countries where agriculture constitutes a large part of the economy; this is especially true for rain-fed agriculture, which is highly correlated with low GDP per capita. Of course, advanced economies also face challenges: looking at US data, Aruoba et al. (2016) show that expenditure-side and income-side GDP estimates, though highly correlated, lead to different growth estimates. Likewise, Deaton (2005, Table 2) compares the average difference between GDP estimates based on national accounts and income estimates based on household surveys across countries, showing that the difference is *smallest* for countries in sub-Saharan Africa—though the coefficient of variation is also *greatest* for countries in sub-Saharan Africa, implying that changes over time are potentially more heavily influenced by measurement error in that region.

In this article, we first investigate the reliability of growth measurement across countries by comparing several data sources. We begin with a brief overview of GDP measurement and a discussion of the measurement challenges faced by all countries. We then triangulate and compare growth estimates based on several different data sources and methods: national accounts, household surveys, and satellite data on night-time light sensors and on vegetation mappings. While each source measures a different concept—so would not be expected to yield identical growth estimates—we interpret a tight concordance between different estimates as a sign of growth estimate reliability. We find that—contrary to common perceptions—there is no compelling evidence that growth is on average measured less well in developing countries. However, we find consistently higher dispersion in growth data for developing countries, which lends support to the view that perceptions about growth (mis)measurement may be due to higher levels of classical measurement error or the existence of a few problematic outliers.

We then turn to several measurement challenges specific to developing countries: limited statistical capacity, the use of outdated data and methods, large agricultural sectors, large informal economies, and limited price data. Using a newly constructed indicator of statistical integrity based on novel IMF audit data, we

do not find compelling evidence that statistical integrity is a first-order issue in most developing countries. We conclude by identifying concrete steps to improve growth measurement in developing countries, including strengthening statistical capacity and supplementing traditional growth measurement approaches with information from innovative data sources. For example, satellite-based vegetation data can measure activities by smallholder farmers that are less likely to be captured in GDP estimates, and several other new data sources offer scope to complement the standard methods. Overall, developing countries (especially low-income countries) perform better than expected at estimating output and growth given the constraints they face, but there is ample room for improvement.

A Brief History of National Income and Growth Measurement

While the notion of measuring economic growth has existed for centuries, today's commonly used methods are typically credited to the work of Simon Kuznets and Richard Stone. In the 1930s, the Great Depression created a desire to measure the severity of the crisis and any progress toward ending it (Kuznets 1934). In a powerful example of economic research informing policy, Kuznets reported on his work to the US Congress, and by 1942, the US government began publishing estimates of gross national product (GNP), in part to aid in war planning efforts. Around the same time, the United Nations (UN) recognized the value of measuring economic progress using methods that were consistent over time and comparable across countries. Stone helped the UN Committee on National Income Statistics develop a framework for a System of National Accounts (SNA) (Stone 1947a), and in 1953 the UN Statistical Commission released SNA guidelines that were applicable for most of the world, including developing or lower-income countries (Stone 1953). Both Kuznets and Stone would eventually receive the Nobel Prize for their work in developing and refining national growth accounting methods: Kuznets in 1971 (just the third Nobel Prize in economics ever awarded) and Stone in 1984 (<https://www.nobelprize.org/prizes/economic-sciences/1971/kuznets/facts/>; <https://www.nobelprize.org/prizes/economic-sciences/1984/stone/facts/>).

Since the original 1953 guidelines on the System of National Accounts, there have been a series of revisions to improve the quality of the measures and address measurement error, overseen by the Inter-Secretariat Working Group on National Accounts (ISWGNA)—a body comprising members from the International Monetary Fund (IMF), the European Union, the Organization for Economic Co-operation and Development, the UN, and the World Bank.¹ For example, following the most recent update to the SNA guidelines in 2008, the ISWGNA developed an *Implementation Programme for the System of National Accounts 2008 and Supporting Statistics* to assist

¹For a history of these revisions, see Figure A1 in the online Appendix available with this article at the *JEP* website, or the UN Statistics Division website at <https://unstats.un.org/unsd/nationalaccount/hsna.asp>.

countries in building the statistical and institutional capacity needed to successfully transition to the new guidelines.

In addition to helping establish the System of National Accounts, Stone wrote seminal papers in the 1940s on measurement error in estimating national income. This early literature leveraged the variations in national income estimates from different measurement approaches (that is, expenditure-side and income-side) to assess and address measurement error (Stone, Champernowne, and Meade 1942). This approach is also the basis of recent literature, including Aruoba et al. (2016). Economists since Kuznets have long been familiar with the basic conceptual criticisms of GDP: that it fails to capture important aspects of well-being like leisure, health, and environmental protection, for example, or that it omits information about the distribution of income (Sen 1985; Nussbaum 1987; Stiglitz, Sen, and Fitoussi 2009).²

Despite concerns over measurement and interpretation, for decades nearly all countries worldwide have reported GDP and used the measure as a critical factor for short- and long-term policymaking.

Are Growth Estimates Less Reliable in Developing Countries?

There is no single, well-defined metric to assess the reliability of a country's national income and growth statistics. The most common approach, similar to Stone, Champernowne, and Meade (1942), is to compare growth estimates obtained using different data sources and approaches to examine whether the estimates coincide or are correlated. In this article, we explore three main conceptual constructs for the estimation of economic growth and make comparisons among them to assess the reliability of growth estimates.

The central measure we examine is GDP per capita, estimated based on System of National Accounts standards and usually produced by each country's National Statistical Office. As taught in introductory economics classes, GDP can be viewed as the sum of personal consumption, investment (including change in inventories), government expenditures, and net exports. Alternatively, it can be viewed as the sum of personal income, tax revenues on production and imports, and corporate tax revenues (including undistributed corporate profits).

We compare this standard measure to two alternative approaches. First, we consider household surveys of budgets, income, expenditure, or consumption. The

²Sen (1985) and Nussbaum (1988) argue that well-being is linked to the capability of an individual to live a life the person has reason to value. This is interpreted as being able to live a healthy life and participate in society without shame. This capabilities approach to measuring well-being underpins the United Nations Human Development Index. The Stiglitz, Sen, and Fitoussi (2009) critique of GDP is twofold. First, GDP fails to account for the within-country distribution of income. Second, some actions increase GDP but reduce well-being (like traffic jams leading to higher fuel consumption and a reduction in well-being), and similarly, some activities contribute to well-being but do not increase GDP (like unpaid household labor).

common method for this approach is to extract per capita household consumption or per capita household income, and then to compare growth rates of these measures with measures of personal income, personal consumption, or GDP per capita growth based on the System of National Accounts. A substantial share of household survey data is collected through large-scale efforts supported by the World Bank, such as the Living Standards Measurement Study. There are many reasons survey-based and SNA-based measures will differ. For instance, SNA protocols for income-side measures place relatively less emphasis on capturing informal economic activities, such as subsistence farming or so-called shadow activities such as the production of illegal drugs. Because household surveys in lower-income countries typically focus on asking people questions about what they have consumed (rather than what they have earned), they are more likely to capture such activities. Another difference is that SNA-based protocols place greater emphasis on larger transactions relative to smaller transactions, which have little impact on total income measures; in fact, Deaton (2005) documents that SNA training instructions directly specify that greater effort should be directed at larger transactions. In contrast, household budget and living standards surveys tend to include regular smaller transactions with greater probability than (often irregular) larger ones like weddings and funerals (Deaton and Zaidi 2002).

Next, we consider an approach that has only become possible in recent years: using satellite data for economic analysis (discussed in this journal by Donaldson and Storeygard 2016). Night-time lights have received particular attention, especially the Defense Meteorological Program (DMSP) Operational Linescan System (DMSP-OLS). Luminosity can serve as a proxy for economic activity, and night lights provide frequent, relatively cheap, and globally available data (Chen and Nordhaus 2011; Henderson, Storeygard, and Weil 2012; Pinkovskiy and Sala-i-Martin 2016).³ Like other measurement approaches, night lights are imperfect. Zhou et al. (2015), for example, argue that limitations in the sensor of these lights create saturation problems in central urban areas, potentially hampering their ability to predict variation in economic activity in rich, high-density areas. By contrast, Gibson et al. (2021) argue that DMSP-OLS light data are poor predictors of economic activity in low-density, rural areas. An additional data source that potentially can be harnessed for growth measurement is satellite-based vegetation indices, estimated using reflectance from plants. A Normalized Difference Vegetation Index (NDVI) is estimated

³Night-time lights data are publicly available in an easy-to-use format from the National Oceanic and Atmosphere Administration (NOAA) website from 1992–2013. The site provides several data series. One frequently used night-time lights series is from the Defense Meteorological Satellite Program–Operational Linescan System (DMSP-OLS). This data source is cleaned to capture luminosity separate from the effects of cloud coverage, fires, aurora, and ephemeral light (Elvidge et al. 2009). Newer sources of night lights, such as the Visible Infrared Imaging Radiometer Suite (VIIRS), have also emerged; however, this data source is less regularly cleaned and is accessible for only a few years.

by satellite detection of reflectance from plants in specific portions of the visible and infrared spectra.⁴

Of course, one would not expect these various data sources to yield identical growth estimates. National accounts, household surveys, and satellite data were each designed for different reasons and serve different purposes. Nevertheless, we would expect the growth rates they generate to be correlated. Accordingly, in the following sections we examine correlations, long-run trajectories, and some key differences across these approaches. In the context of these comparisons, we examine whether growth-estimate reliability varies by country income grouping. In instances where such comparisons exist from earlier studies, we update them to more recent years and extend them to more countries.

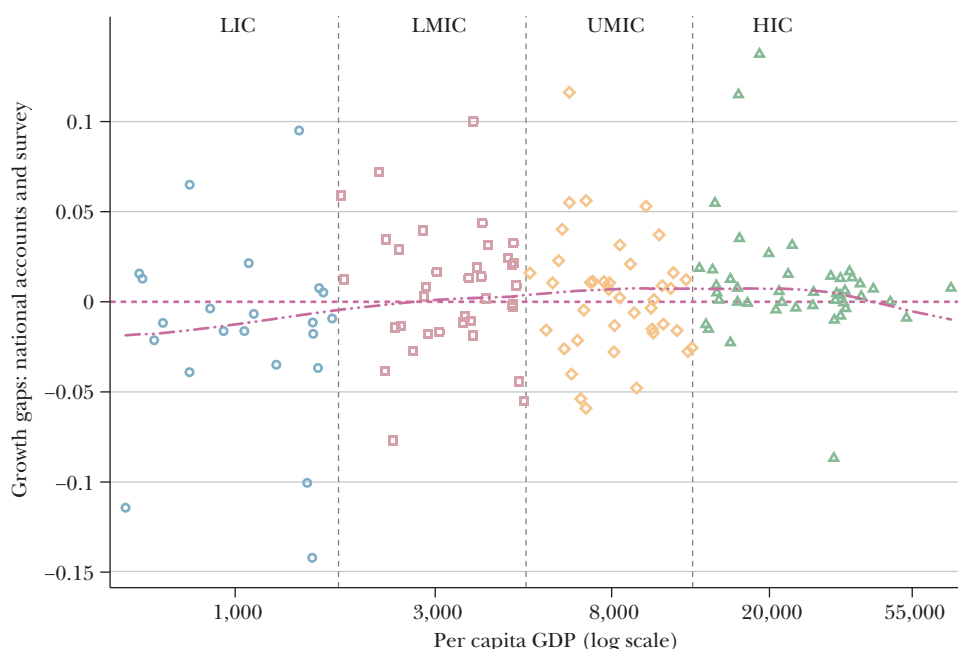
Growth Estimates from National Accounts and Household Surveys

It is well established that there are significant gaps between national accounts estimates of GDP or personal consumption and household survey estimates of income or consumption (Deaton 2005; Ravallion 2003). Prydz, Jolliffe, and Serajuddin (2020) updated this earlier analysis by examining the ratio in levels of GDP (and household final consumption expenditure) to income (and consumption) from a series of more recent household surveys, finding that middle-income—not low-income—countries have the weakest relationship between national accounts and survey measures. A potential explanation for this finding is that middle-income countries often have fast growth, which could decrease survey response rates (as households become richer, the opportunity cost of their time increases) and produce a downward bias in survey-based growth estimates. In addition, a more rapidly changing economic structure might increase discrepancies in income measurement if, for example, national accounts do not adjust the weights of industries that have become increasingly important over time (as was the case in Nigeria’s 2014 GDP rebasing, for example). Broadly speaking, in the literature regarding GDP level estimates, there is an unresolved debate regarding the reliability of national accounts data by country income grouping, with conclusions varying among the leading studies.

Rather than examining levels, which have been examined in earlier papers and which are expected to differ across data sources given that different data measure different concepts, we focus our comparisons on growth rates. Growth-rate comparisons are, in principle, subject to the same caveats regarding differences in concepts measured, but we expect these caveats to be less consequential given that the focus on growth rates controls for the impact of time-invariant differences across measures. Also, annual growth rates often receive the most coverage and attention

⁴We use a data series for 89 economies where over 25 percent of employment is in agriculture from 2000 to 2018. We include measures for total Normalized Difference Vegetation Index per year per country as well as the maximum versus minimum NDVI in a given year and country. Based on definitions from NASA’s Moderate Resolution Imaging Spectroradiometer (MODIS), we also disaggregate the NDVI by smallholder farms, which are often part of the informal economy versus large-scale commercial agricultural land, which usually is captured in national accounts.

Figure 1

Gaps in Average Growth between National Accounts and Survey Estimates

Source: Author calculations using data from the World Bank.

Note: A positive gap indicates GDP growth rates are higher than household surveys and vice-versa. Growth estimates are averaged over the time period 1992–2012 and are not weighted by the population of each country. Each income category is labeled: LIC = Low-income country; LMIC = Lower-middle-income country; UMIC = Upper-middle-income country; HIC = High-income country.

in international policy dialogues. However, measures of year-on-year growth are often volatile, and their variations are potentially due to noise. To minimize the impact of noise on our comparisons, we average annual growth rates over the time period 1992–2012 for each country.

Figure 1 provides a comparison of average growth rates based on estimated GDP per capita from national accounts and estimated per capita income or consumption from household surveys. For the household survey measure, we extract data from the World Bank's PovcalNet which provides a mix of per capita household consumption and income measures, depending on what is available (at <http://iresearch.worldbank.org/PovcalNet/home.aspx>).⁵ For each country, we estimate

⁵The majority of these data files are based on integrated household surveys such as those in the Living Standards Measurement Study. The consumption aggregate is a broad measure, which includes consumption of food and nonfood items, with food consumption including food purchased from the market, home-produced food, and payment-in-kind. Nonfood consumption typically includes the total value of

the average growth rate over 1992–2012. The figure plots the gap between the two by income grouping. The most notable feature is the dispersion by income category, which is visibly highest for low-income countries and lowest for high-income countries. While the gap between different growth estimates is not significantly higher for developing countries on average, it is very large for select low-income countries.

As with the literature on GDP levels, the comparison of growth rates based on national accounts and survey data does not offer clear-cut conclusions. Our results seem consistent with the view that growth measurement may be most problematic in low-income countries, though as noted earlier, this view is supported more by the high dispersion of growth estimate gaps in low-income countries than by the size of the average gap.

GDP, Household Surveys, and Night-Time Lights Data

Next, we add into the analysis average growth rates based on satellite-based night light data by country from 1992 to 2012. While night light data have limitations and are an imperfect proxy for economic activity, they have two notable advantages: they are not biased by potential non-response, as household surveys are, and they are not easily manipulated or frequently adjusted, as national accounts data might be. Figure 2 plots a smoothed nonparametric regression of the growth rate based on each measure on log GDP per capita.

We observe a few patterns in the data. First, the GDP line shows the growth rate of per capita GDP over the range of countries based on national accounts data, which shows that middle-income countries grow more quickly than either high-income or low-income countries.

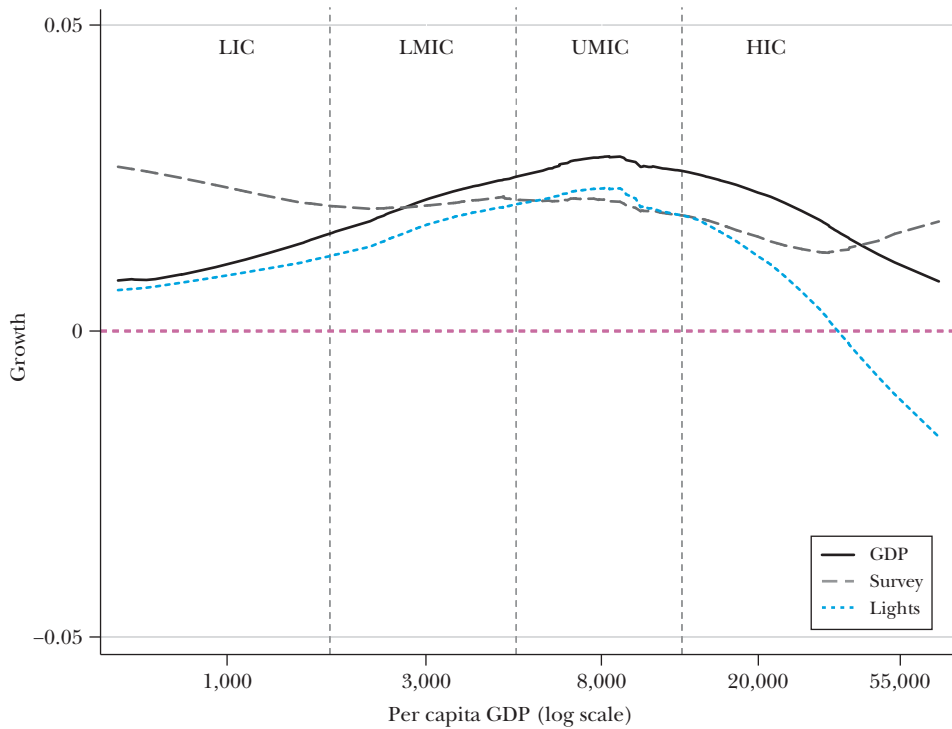
Second, growth rate estimates based on survey data are lower than estimates based on national accounts data for all categories *except* low-income countries, where survey estimates are higher: survey-based growth is on average 2.6 percent while national accounts-based growth is slightly less than 1 percent. One reason for this pattern might be that survey estimates capture more informal economic activity, which comprises a large share of the economies of low-income countries and which national accounts estimates may be less suited to measuring.

Third, light growth tracks GDP growth closely in all income categories except high-income countries. This suggests that lights might be useful in triangulating accurate GDP estimates in developing countries but that the relationship might be less clear for high-income countries. This pattern has several potential explanations: growth measurement could be less reliable in high-income countries; urban saturation in high-income countries might dampen light growth estimates; or the relationship between lights and economic growth (as measured by either national accounts or household surveys) could be non-monotonic by income level. For

small nonfood items plus the use-value of durable goods. For high-income countries, the majority of the PovcalNet data comes originally from either Eurostat's Statistics on Income and Living Conditions or the Luxembourg Income Study, which creates an income vector that is harmonized across countries in their archives.

Figure 2

Average Growth across Measures: GDP, Survey, and Lights



Source: Author calculations using data from the World Bank and NOAA.

Note: We run a LOWESS smoothed nonparametric regression of growth rates by income level on log GDP per capita terms. Each income category is labeled: LIC = Low-income country; LMIC = Lower-middle-income country; UMIC = Upper-middle-income country; HIC = High-income country. The categorization of countries is based on the current World Bank classification. For details of the calculations, including average growth rates for each measure and standard deviations across countries and over time, see the online Appendix.

example, some high-income countries have tried to reduce light pollution, in which case light would have a negative rather than positive association with income.

Fourth, while the gap between survey and national accounts is largest in low-income countries, for lights the gap is smallest in low-income countries. Hence, whether one considers growth measurement to be more or less reliable in developing countries may depend on which alternative measure one trusts most: lights or surveys.

However, if we examine variation across countries, we find a more consistent pattern, with high variation among low- and middle-income countries in the gaps between national accounts and surveys as well as lights. In Table 1, cross-country variation in national accounts growth estimates ranges from 2.2 to 3.8 percent

Table 1

Average Growth across Measures—GDP, Survey, and Lights

	<i>Low income</i>	<i>Lower middle income</i>	<i>Upper-middle income</i>	<i>High income</i>
<i>Growth GDP</i>				
Mean	0.009	0.025	0.028	0.023
Across country SD	0.038	0.022	0.031	0.018
Within country SD	0.052	0.044	0.050	0.030
<i>Growth survey</i>				
Mean	0.026	0.016	0.026	0.013
Across country SD	0.047	0.032	0.027	0.031
Within country SD	0.055	0.062	0.084	0.059
<i>Growth lights</i>				
Mean	0.009	0.017	0.025	0.011
Across country SD	0.028	0.032	0.047	0.023
Within country SD	0.183	0.152	0.159	0.196
Observations	25	38	39	40

Note: Growth rates are calculated as the log first difference. We average growth rates per country from 1992 to 2012 for each measure to account for year-to-year noise and variation. We then average country average growth rates by income category. Averages are not weighted by the population of each country. “Across Country SD” refers to the standard deviation of growth rates across countries, averaged by income category. “Within Country SD” refers to the standard deviation of growth rates *over time* within a country, averaged by income category.

in low- and middle-income countries, respectively, relative to 1.8 percent in high-income countries. Similarly, we observe higher cross-country variation in developing countries for survey and lights data. This points to a potential “black sheep” explanation: while discrepancies in growth estimates are not systematically worse in developing countries on average, there are a few countries for which such discrepancies are particularly large, and these cases may be responsible for the perception that growth measurement in developing countries is unreliable.

We observe a similar pattern for within-country variation of GDP growth estimates based on System of National Accounts data over time, ranging from 4.4 percent to 5.2 percent in low- and middle-income countries, compared to 3 percent in high-income countries. Again, this evidence suggests variation and volatility might play an important role in perceived reliability of growth estimates in developing countries.

Finally, we examine the role of limited data availability in some countries. We find that within-country correlations of survey and national accounts over time are higher at higher-income levels, varying from 0.16 to 0.33. Table A1 in the online Appendix, available with this article at the *JEP* website, shows within-country, year-to-year correlations between measures. However, this pattern virtually disappears when restricting the sample to countries with survey data for more than three time periods. Hence, it seems that the lower year-to-year correlations in

low-income countries are driven by limited data. For example, Rwanda has only three survey data points, meaning that growth rates can only be estimated at two points in time, and any correlation in estimates between survey and national accounts over time is derived from the single difference in growth from 2005 to 2010. As another example, Figure A2 in the online Appendix shows only five household survey data points in Tanzania between 1992 and 2012, relative to over 20 in Indonesia.

Summary

While some statistics suggest less reliable growth measurement in developing countries, the cumulative evidence is mixed. Previous work exploring correlations in GDP levels has not found evidence that low-income countries underperform higher-income countries in measurement. Similarly, we do not find systematic evidence based on night lights data that growth is measured less well on average or manipulated in developing countries. Light estimates in low-income countries follow a similar trajectory as national accounts estimates, and if anything, they track each other more closely than in high-income countries. In general, different comparisons lead to different conclusions. These results reinforce the value of supplementing national accounts estimates with survey-based measurement and of utilizing alternative sources of income estimates, such as satellite data, as we discuss in the paper's final section.

However, a consistent finding across all comparisons is that cross-country dispersion in growth estimates is substantially higher in developing countries, suggesting a possible role for a few outliers to generate the perception that all developing countries' growth estimates cannot be trusted.

Finally, we note that differences in average growth rates across the three different measurement approaches appear small—typically around 1.5 percentage points or less. While gaps of this magnitude may be considered large for high-income countries, where annual growth rates have recently been in the 3–4 percent range, they appear small for many fast-growing developing countries. We conclude that even though growth estimates may be imprecise, they are likely trustworthy within a margin of error of about 1.5 percentage points. Considering the uncertainty around such estimates, this error margin does not seem grave. It also suggests that paying excessive attention to potentially noisy year-on-year growth estimates seems unnecessary. At a minimum, year-on-year growth estimates should be accompanied by confidence intervals, which should be given as much attention as the estimates themselves.

Measurement Challenges in Developing Countries

While measurement challenges exist for both developed and developing countries, in this section we turn to the specific challenges that developing countries face in estimating growth.

Low Statistical Capacity and Lack of Independence of Statistical Authorities

The term “developing” signifies vulnerabilities and resource constraints that affect many areas, including data collection and production of statistics (Carletto, Jolliffe, and Banerjee 2015; Devarajan 2013; Jerven and Johnston 2015). Many developing countries use old data, outdated methods, and unreliable statistics due to lack of funding, inadequate resources for data collection, management and dissemination, and absence of coordination among relevant agencies and stakeholders. Statistical capacity constraints are particularly relevant in Africa (Devarajan 2013). As of early 2021, only about one-third of sub-Saharan African countries use the most recent System of National Accounts standards from 2008, while most of the rest use the 1993 standards.

Changing from one vintage of the System of National Accounts to another, or infrequent updates to growth accounting methods, can lead to substantial GDP movements, which in turn may contribute to the perception of unreliable or manipulated growth measurement in developing countries. For example, Ghana’s adoption of the 1993 SNA system in 2010 led to a 62 percent upward revision of GDP (Devarajan 2013), and Ghana has since adopted the 2008 SNA. A similar revision in Malawi led to a 32 percent upward GDP revision. Likewise, failing to regularly update the base year for GDP estimation, which determines the weights reflecting the relative importance of different sectors, can create discrete breaks in a country’s GDP series.⁶ In addition to the aforementioned case in Nigeria, other examples include Senegal’s 2014 rebasing (from 1999), which increased GDP by 29 percent, and Zimbabwe’s 2012 rebasing (from 2009), which increased GDP by 20 percent. For a systematic view of countries’ statistical capacity, in 2004 the World Bank developed the Statistical Capacity Indicator (SCI). Scores range from 0 (no statistical capacity) to 100 (adequate statistical capacity), with an overall score as well as scores in three sub-categories: Source Data, Methodology, and Periodicity.⁷ The SCI’s source data are collected mostly for low- and middle-income countries.

The average Statistical Capacity Indicator score for a low-income country is about 60, which is similar to the average regional score for the Sub-Saharan Africa and Middle East & North Africa regions. Lower- and upper-middle income countries have an average score of about 70, which is similar to the average regional score for Latin America & Caribbean and South Asia. Several low- and middle-income

⁶The US Department of Commerce’s Bureau of Economic Analysis (BEA) introduced chain-weighting specifically to overcome this problem of discrete changes in GDP trends from occasional updates to a fixed base year (Steindel 1995).

⁷The source data for the Statistical Capacity Indicators refers to surveys for agriculture, health, poverty, the population census, and vital registration systems. The Methodology sub-category considers the following components: balance of payments manuals, consumer price index base year, external debt reporting status, government finance accounting, import and export price indices, industrial production price indices, national accounts base year, national immunization coverage, special data dissemination standards, and UNESCO reporting. The Periodicity sub-category refers to regular data for multiple categories including education, health, sanitation, and gender equality as well as GDP. The SCI Dashboard provides information on the time series of SCI, so that one can track countries’ progress towards statistical capacity (<http://datatopics.worldbank.org/statisticalcapacity/SCIdashboard.aspx>).

countries, such as South Africa and India, score well on statistical performance, indicating that lower income is not synonymous with bad data.⁸ The World Bank recently released the Statistical Performance Indicators (SPI), an update and re-conceptualization of the SCI.

As an alternative way to assess countries' statistical capacity and data quality, the IMF recently released a rich new dataset with information gathered in the process of compiling growth statistics (Berry et al. 2018). In contrast to the Statistical Capacity Indicator scores, which includes multiple statistics not directly linked to growth (such as education statistics reported to UNESCO), the IMF data focuses exclusively on data behind the System of National Accounts and also includes high-income countries. We observe some notable trends. First, the average SNA vintage is consistently older in low-income countries, aligned more closely to the 1993 guidelines than the more recent 2008 vintage commonly used in high-income countries. In addition, the GDP base year is older in low- and middle-income countries, which, as noted, increases the likelihood that national accounts will fail to reflect important changes in a country's economic structure (while also increasing the likelihood of large and potentially contested GDP expansions when the base year is ultimately updated). Second, while "availability of annual GDP" is similar across income categories, "availability of quarterly GDP" estimates varies substantially by income level, ranging from 38 percent for low-income countries to 91 percent for high-income countries. Third, the share of countries that independently compile GDP using different approaches (for example, based on expenditure and on production), which can enhance the reliability and quality of national accounts statistics, also varies by income level: 12 percent for low-income countries, 30 percent for lower-middle-income countries, 40 percent for upper-middle income countries, and 76 percent for high-income countries. A variety of other indicators are available in the IMF data, including timely release of annual or quarterly GDP data and advance release calendars.⁹

Here, we present a novel database of indicators based on expert audits of national accounts called the Reports on the Observance of Standards and Codes, which is a large initiative carried out jointly by the World Bank and the IMF to monitor compliance with international standards for statistical systems (for details, see <https://www.imf.org/en/Publications/roscc>). These reports assess criteria of the IMF Data Quality Assessment Framework (DQAF) for 83 countries. A main advantage of this new database is that it identifies additional quality measures that go beyond a focus on GDP compilation practices: as one example, there is an indicator related to revision policy and practice, which are viewed by the IMF as central to data quality. Each indicator is assessed by IMF auditors based on four rankings:

⁸For details of these calculations, along with a map showing these patterns by country, see the online Appendix available with this paper at the *JEP* website, especially Figure A3 and Table A2.

⁹The summary statistics presented here are compiled and structured from text responses to periodic IMF surveys conducted with 189 countries globally. We average statistics by income category. For detailed tables, see the online Appendix available at the *JEP* website.

Table 2

Summary Statistics for Systems of National Accounts: Capacity, Quality, and Integrity

	<i>Quality</i>		<i>Capacity</i>		<i>Integrity</i>		
	<i>Revision</i>	<i>Monitoring and process</i>	<i>Data use</i>	<i>Resources</i>	<i>Statistical professional practice</i>	<i>No prior data access</i>	<i>Legal environment</i>
High income	0.92	1.00	1.00	0.88	1.00	1.00	0.96
Upper-middle income	0.96	0.96	1.00	0.71	0.96	0.83	0.92
Lower-middle income	0.95	1.00	0.75	0.65	0.95	0.95	0.85
Low income	0.80	0.90	0.60	0.30	1.00	0.90	1.00
East Asia & Pacific	0.86	1.00	1.00	0.71	1.00	1.00	1.00
Europe & Central Asia	0.97	0.97	1.00	0.80	0.97	0.97	0.97
Latin America & Caribbean	0.93	1.00	0.93	0.67	1.00	0.93	1.00
Middle East & North Africa	0.86	1.00	0.86	0.86	0.86	0.86	0.86
North America	1.00	1.00	1.00	1.00	1.00	1.00	1.00
South Asia	1.00	1.00	0.50	0.75	1.00	1.00	0.50
Sub-Saharan Africa	0.87	0.93	0.67	0.40	1.00	0.80	0.87

Note: This table summarizes novel data compiled by the World Bank and IMF and aligned to the United Nations Fundamental Principles of Official Statistics. IMF staff routinely conduct in depth audits with countries around the world including visits to National Statistics Offices and joint review of data sources and process documentation. We group a subset of the indicators arising from these audits displayed in the left-hand column of online Appendix Table B1, available at the *JEP* website, to three high-level categories: Quality (indicators 4.3 and 0.4); Capacity (indicators 5.1 and 0.2); and Integrity (indicators 1.1, 1.2, and 0.1). Table B1 in the online Appendix includes more background on each indicator.

observed, largely observed, largely not observed, or not observed. For our purposes, we code analysis as a dummy variable equal to one if the practice is observed or largely observed, and zero otherwise.

Table 2 breaks down seven indicators from this new database by income group and region. The first two columns show measures of quality: whether revisions and updates of GDP estimates follow a regular and transparent schedule and whether they are monitored and accompanied by explanatory notes. Low-income countries appear to have lower-quality statistics, which is consistent with the indicators of statistical capacity already presented. For example, 80 percent of low-income countries have revision policies and practices, compared to 92 percent and 96 percent in high- and middle-income countries, respectively.

The next two columns show measures of statistical capacity. We first examine human resources in national statistical offices. While 88 percent of national statistical offices in high-income countries are deemed to have enough human resources, this indicator falls to only 30 percent in low-income countries.

The final three columns seek to measure the potential for politically motivated data manipulation, referred to as data integrity. A surprising pattern in this category

is that the lowest scores are observed in middle-income countries. Only 83 percent of upper-middle income countries specify that there is no internal governmental access to statistics prior to their release. Moreover, only 85 percent of lower-middle income countries have a legal environment that clearly delineates responsibilities for the collection and processing of data, compared to 96 percent of high-income countries and 100 percent of low-income countries. These patterns suggest that manipulation may be more feasible where there exists a threshold level of statistical capacity and sophistication that can potentially be used to promote political agendas.

Overall, constraints on statistical capacity emerge as a major factor affecting the quality of implementing the System of National Accounts in low-income countries, while conditions for deliberate data manipulation are more likely to be observed in middle-income than low-income countries.

The Role of the Agricultural Sector

The agricultural sector contributes about 5 percent of total world economic production but represents a much larger share in most developing countries.¹⁰ In Africa, agriculture is the largest sector and accounts for 15 percent of total GDP. In some developing countries, especially in Africa and South Asia, agriculture represents more than half of economic output (according to the World Development Indicators, <https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS?end=2019&start=1960&view=chart>). In addition, agriculture's contribution to growth volatility is about three times greater than the service sector's contribution (Koren and Tenreyro 2007).

However, agricultural production is often poorly measured (Jerven and Johnston 2015; Carletto, Jolliffe, and Banerjee 2015). In many low- and middle-income countries, the quantity of crops harvested on cultivated land or the amount of land cultivated are estimated in part through self-reported farmholder surveys which suffer from significant levels of measurement error (Abay et al. 2019; Dillon et al. 2018; Gourley, Kilic, and Lobell 2019). For example, Carletto, Savastano, and Zezza (2013) show that self-reported plot sizes by the bottom decile of farmers (in terms of landholdings) are double what satellite measurements indicate. Similarly, the data used by Desiere and Jolliffe (2018) indicate that self-reported crop yields by the bottom quartile of farmers (in terms of landholdings) are about twice as large as actual yields.

Self-reports of the quantity and value of production are also fraught with measurement concerns. Many subsistence farmers sell relatively little of their crop output but are nonetheless frequently asked to report its value and quantity. When market transactions do inform responses, they are frequently based

¹⁰The CIA World Factbook estimates agriculture value added to be 6.4 percent while the World Bank's World Development Indicators estimates the value added of agriculture to global GDP to be about 4 percent (<https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS>). Figure A4 in the online Appendix available at the *JEP* website shows worldwide estimates.

in nonstandard units such as heaps, piles, buckets, or bags, which are often not comparable beyond a limited geographic area. For example, Capéau and Dercon (2006) note that a *tassa* (serving can) is commonly used to report market transactions in Ethiopia, but the unit of measurement is known to be significantly larger in northern Ethiopia.

For countries with large agricultural sectors, the reliability of estimated GDP growth depends on how well agricultural activities are accounted for in national accounts. As national accounts focus on measuring total output, the methodological approach places greater emphasis on accurately capturing large farms' production. Agricultural household surveys, by contrast, typically focus on understanding constraints to improving yields and profits for smallholder farms, which comprise a sizable share of agricultural activity. Lowder, Scoet, and Raney (2016) estimate that there are 570 million farms worldwide, over 87 percent of which are small (less than 2 hectares or about 5 acres) and family operated. Moreover, 95 percent of smallholder farms are in low- or middle-income countries.

We assess whether the high prevalence of smallholder farms in low- and middle-income countries reduces the reliability of GDP growth estimates. To do so, we compare GDP value-added agricultural growth in national accounts and agricultural growth proxied by a satellite-based vegetation index (see online Appendix available at the *JEP* website for a data description) from 2000 to 2018 across 87 countries.

Table A4 in the appendix shows regression results. We find a positive and statistically significant relationship of .317 for all farms, which suggests that the vegetation index is highly correlated with national accounts estimates of agricultural output. We next disaggregate the vegetation index by smallholder (column 2) and larger corporate agricultural growth (column 3), based on definitions from the NASA Moderate Resolution Imaging Spectroradiometer (MODIS). While large corporate agricultural activity has an even stronger relationship with GDP estimates, reflected in a statistically significant coefficient of .388, smallholder growth has *no statistically significant relationship* with GDP estimates. Including country and time fixed effects (columns 4 and 5) leads to qualitatively similar results.

These results are visualized in Figure 3, which shows the positive relationship between agricultural output based on national accounts estimates and the vegetation index for all farms. This is driven by the positive relationship with large farm output, while there is a strikingly flat and slightly negative relationship with small farm output.

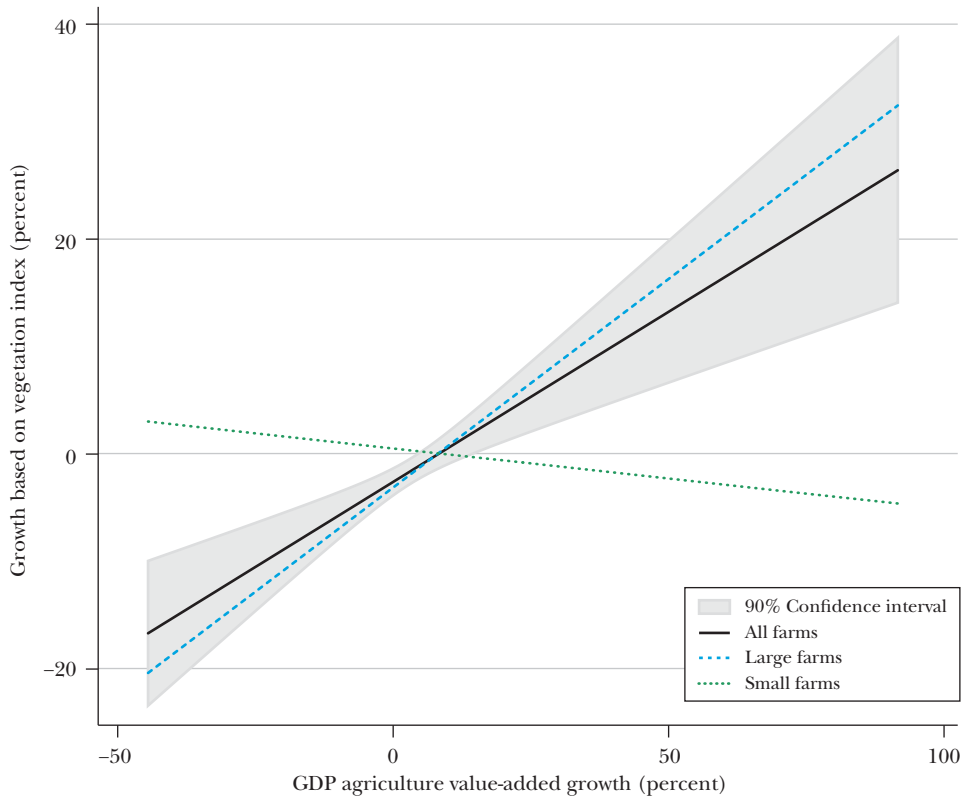
These results are consistent with the interpretation that smallholder agricultural activity is not well captured in official GDP as estimated in national accounts. This has substantial implications for the accuracy of growth measurement in developing economies, where smallholder farms are particularly important.

Informal or Shadow/Underground Economy

Developing countries are also characterized by a large informal sector, defined broadly as economic activity that is invisible to government, either because firms are

Figure 3

Measuring Agricultural Output: Comparing Growth Based on National Accounts and Satellite-Based Vegetation Growth



Source: Author calculations based on satellite images and data from Landstat8, and farm type classifications based on NASA definitions.

Note: Data for the agriculture vegetation index was produced from satellite imagery. The distinction between smallholder and larger corporate farms is based on definitions from the NASA Moderate Resolution Imaging Spectroradiometer (MODIS). For a description of the data, see the online Appendix available at the *JEP* website. We run regressions using panel data from 2000 to 2018 across 87 countries in which the share of agricultural employment is above 25 percent.

not registered (and hence avoid taxes and regulations) or workers are not registered (and hence do not receive social protection). The concept of “informality” was born in Africa; in Ghana, “informal income opportunities” are common for individuals, and in Kenya, it is typical for enterprises to be informal (Charmes 2012).

The informal economy is also occasionally referred to as the “shadow” or “underground” economy. Illegal activities are typically not well captured in measures of GDP based on the System of National Accounts, though they are arguably important in some countries. For example, in Afghanistan, the drug industry is estimated

to comprise as much as one-third of GDP but is largely not accounted for in official growth statistics (Buddenberg and Byrd 2006). In contrast, farmer self-reports of poppy production in Afghanistan's national household survey are substantial and do not appear to suffer from significant nonresponse problems.¹¹

The informal sector represents a major measurement challenge in developing countries—especially in sub-Saharan Africa—for the same reasons that agriculture is a challenge. This is in part because there is substantial overlap between agriculture and the informal economy. Based on data from household surveys in 69 countries, the International Labour Organization (2018) estimates that the informal economy represents 41 percent of GDP in sub-Saharan Africa, ranging from less than 30 percent in South Africa to 60 percent in Nigeria, Tanzania, and Zimbabwe.¹² Charmes (2012) reports that in the 2000s in sub-Saharan Africa, the informal sector (including the agricultural household sector) contributed nearly two-thirds of GDP, with the highest share in Niger (72.6 percent) and the lowest in Senegal (51.5 percent). Excluding agriculture, the informal sector represents approximately one-third of sub-Saharan Africa's GDP. In India, Charmes (2012) estimates that the informal sector comprises 54.2 percent of GDP (or 38.4 percent if agriculture is excluded). Using data from 158 countries from 1991 to 2015, Medina and Schneider (2018) estimate the average size of the “shadow” economy to be 31.9 percent of GDP, with the highest shares in Zimbabwe (60.6 percent) and Bolivia (62.3 percent). In sum, while specific estimates vary, existing work indicates that the share of the informal economy in low-income countries is substantial.

The contribution of informal enterprises to GDP can be measured in multiple ways, including surveys of establishments and households or by the residual difference between national expenditure and income statistics. However, since the contribution of informal labor employed in formal enterprises (as an intermediate input) is not included in GDP measurement of final output, this approach likely results in underestimates of the contribution of the informal sector to GDP. Moreover, it also likely results in underestimates of growth, as it is generally believed that informal employment in formal enterprises is growing in developing countries. This underestimation is more pronounced in countries with both large informal employment and a large number of formal enterprises, which tend to be middle-income economies.¹³ Overall, these considerations suggest that the mismeasurement of the

¹¹ Buddenberg and Byrd (2006) provide several explanations for this, one of which is that there is a tradition of openness about discussing poppy production in part due to the legacy from when opium bazaars were common and out in the open. There is also the issue that the household interview is about crop production and not drug production, and that the poppy is just one crop of many that the farm households are asked about.

¹² For estimates of the informal economy worldwide, see Figure A5 in the online Appendix available at the *JEP* website.

¹³ On the other hand, as Charmes (2012) points out, the way the informal sector is treated in the SNA-based measures of GDP may also lead to overestimates of its contribution to GDP because current measurement practice is premised on the assumption that the household sector can be assimilated into the informal sector. This assumption may be true in low-income countries characterized by subsistence

informal sector's contribution to growth may be a bigger issue in middle-income than low-income countries.

These measurement challenges are presumably biggest during policy changes that affect the formal and informal sectors differently. For example, in India real gross value-added growth for the informal sector is proxied by the Index of Industrial Production, which is mostly composed of formal sector firms. While this approach may work reasonably well during normal times, it likely overstated growth in the aftermath of India's demonetization and the Goods and Services Tax (GST)—both policy changes that have been shown to have disproportionately impacted the informal sector (Subramanian 2019; Chodorow-Reich et al. 2018).

Price Measurement

Price deflators are needed to obtain changes in real GDP, but prices are often poorly measured in developing countries. For example, a recent controversy in Rwanda regarding poverty measurement resulted from differences in inflation measurements: while the Consumer Price Index (CPI) suggested that Rwanda's inflation rate from 2011 to 2014 was 23 percent, the National Institute of Statistics in Rwanda (NISR) used a 4.7 percent inflation rate to calculate poverty rates. There were also substantial differences in inflation rates between urban and rural areas, which are largely not captured in the official price index (as reported by Wilson and Blood 2019). In India, Subramanian (2019) flags that the use of a manufacturing Wholesale Price Index as a proxy for producer prices of services in the mid-2010s, a time of sharply declining oil prices, could have led to gross value-added and real growth being overstated.

Some prominent data series on national income lack underlying data on price levels, particularly in developing countries. Young (2012) notes that in 2006 the UN National Accounts database providing GDP estimates in current and constant prices was missing more than half of all 1,410 observations across 47 sub-Saharan African countries from 1991 to 2004. Moreover, among 15 of the countries for which the complete time series are published, there was no constant price data. Similarly, Young (2012) notes that the purchasing power parity index in the Penn World Tables (PWT) version 6.1 provides real incomes for 45 sub-Saharan African countries, but 24 do not have a benchmark study of prices. In 2005, the World Bank's International Comparison Program (ICP) measured prices for 146 countries, for the first time including many previously-excluded developing countries. Accordingly, a substantial revision was conducted between PWT 6.1 and PWT 7.0 to include this new price data, resulting in large differences between countries in per capita income and larger growth estimates for many countries, especially in Africa (Young 2012).

The IMF has collected a dataset to assess statistical practices for price indexes in 193 economies along a variety of dimensions (Berry et al. 2019). The data show

agriculture and a small formal sector, but is less likely to hold in emerging economies with larger and growing formal sectors.

that while consumer price indexes are available in all economies, compliance with the international standard Classification of Individual Consumption According to Purpose (COICOP) system varies substantially by income category: there is 92 percent adoption in high-income countries, and between 57 and 75 percent adoption in low- and middle-income countries. While 87 percent of high-income countries have national expenditure coverage in their consumer price index, only 62 percent do so in low-income countries, with a substantial share (25 percent) of countries deriving price information from capital cities only.¹⁴

When looking at information on producer price indices in this IMF data, we find a steep gradient of data availability by income category. Availability of information on producer price indexes is 79 percent for high-income countries and only 41 percent for low-income countries. The timeliness of producer price index data follows a similar pattern, with monthly data available for 63 percent of high-income economies but only 12 percent of low-income economies. In terms of alignment to a recent classification system vintage, such as to the International Standard Industrial Classification of All Economic Activities (ISIC) Revision 4, 56 percent of high-income economies align with this vintage, compared to less than 9 percent of low-income economies. In addition, when developing a producer price index, the IMF recommends starting with the mining, manufacturing, and utilities sectors, and expanding coverage to more complex activities, such as services over time. We find that while around 60 percent of high-income countries include at least the mining, manufacturing, and utilities sector, only 16 percent of low-income economies do so. No low-income country includes any sectors beyond mining, manufacturing, and utilities.¹⁵

Finally, we examine the practice of “inflation targeting” which refers to the central banking policy aimed at achieving a specific annual rate of inflation. This practice is seen to be a strong proxy for the quality of national accounts systems as it provides a direct incentive for national statistics offices and government ministries to have accurate and timely price information (Carson, Enoch, and Dziobek 2002). We find that while 65 percent of high-income countries practice inflation targeting, only 12 percent of low-income countries do. This indicator can be viewed as a summary statistic for many of the more specific price indicators, as quality and timeliness of each of the specific price indicators makes this practice possible.

Altogether, we find strong evidence from the IMF data that price data in lower income countries is often lacking, out-of-date, or not aligned to best practice.

¹⁴For detailed data on price indexes by country and the Classification of Individual Consumption According to Purpose system, see Table A5 in the online Appendix available with this paper at the *JEP* website.

¹⁵For a table showing a more detailed list of price index practices compiled by the IMF across 193 economies by Berry et al. (2019) as well as a breakdown by high-income, upper-middle income, lower-middle income, and low-income countries, see the online Appendix available with the paper at the *JEP* website.

Do Measurement Challenges Explain Gaps between Alternative Growth Measures?

We now examine how each factor discussed in the previous subsections influences the reliability of growth measurement, as proxied by concordance among various growth measures. For example, when we compare GDP and lights data, the average elasticity between the growth estimates based on these two data sources is around 0.37. If a country's GDP is substantially higher than the best-fit line, this raises a flag that the country might be manipulating its GDP estimates; GDP can be manipulated to higher numbers for political purposes but satellite-based night-lights data cannot.

Figure 4 shows in Panel A the average elasticity between lights and GDP, illustrating deviations from the average elasticity for a select group of countries: China, India, Rwanda, Nigeria, Liberia, and Cambodia. Notably, the GDP growth estimates for China, India, Nigeria, and Rwanda—each of which have faced controversies regarding their statistics—lie *above* the line of best fit, which is consistent with the idea that these countries might be reporting higher growth relative to real economic activity for political purposes. However, this divergence could also be due to other factors; for example, using an inaccurate price index to calculate real GDP could inflate GDP relative to real economic activity. In the case of Cambodia and Liberia, which lie *below* the line of best fit, the divergence might be explained by the countries' large informal economies, which can be observed by night lights but are not fully accounted for in GDP estimates.

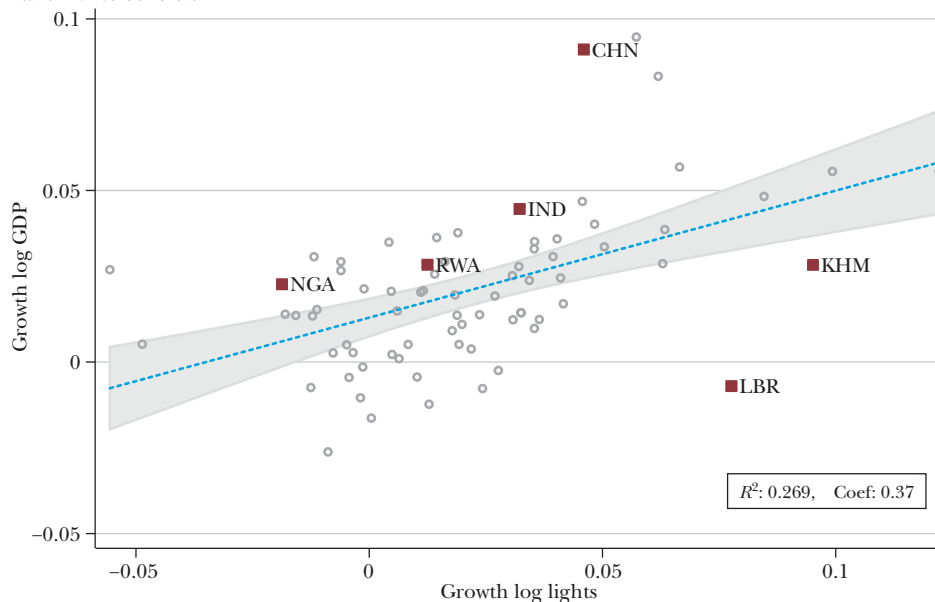
We examine whether controlling for factors that we suspect may be responsible for growth mismeasurement reduces the divergence from the average elasticity and increases the *R*-squared of the associated regression. In Figure 4, we focus on a subset of 74 countries which heavily rely on agriculture (defined as a share of employment in agriculture that is over 25 percent). The unconditioned correlation in Panel A between the log of growth in GDP and night lights suggests a series of countries have growth rates that differ substantially from what is predicted by lights data. Controlling for a series of other indicators, including the satellite-based vegetation index (which plausibly captures smallholder agricultural economic activity) as well as agricultural value-added in national accounts and price measurement practices (Panel B), results in a tighter concentration around the fitted line as revealed by the substantial increase in *R*-squared from 0.269 to 0.577. Several countries (for example, Cambodia, India, Liberia, Nigeria, and Rwanda) are no longer outliers. This suggests that the divergence observed in Panel A may have been driven by the presence of smallholder agriculture, the informal economy, and challenges in measuring price changes. Notably, China does not converge substantially, suggesting the plausibility of GDP data manipulation.

In the online Appendix, available with this paper at the *JEP* website, we conduct this exercise with 164 countries (see Figure A6), successively adding more control variables which helps further explain the difference between night-lights data and the SNA-based measures of GDP. For example, when we condition on our IMF data that is based on the Reports on the Observance of Standards and Codes, we no longer observe any outliers among the 60 countries for which we have data (Figures

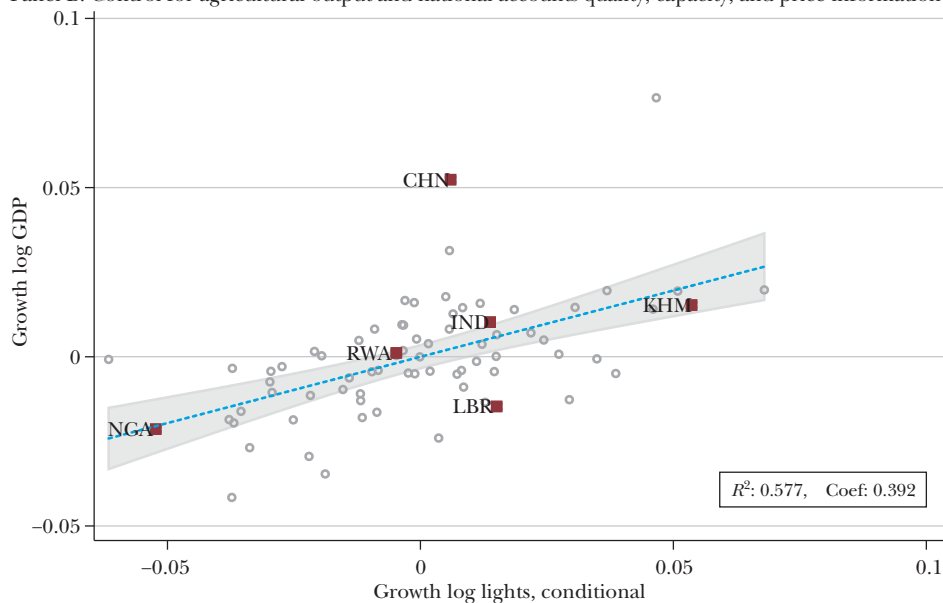
Figure 4

Comparing GDP and Lights with and without Vegetation Index Controls

Panel A. No controls



Panel B. Control for agricultural output and national accounts quality, capacity, and price information



Source: Author calculations based on data from the World Bank, vegetation satellite data from Landstat8, as well as quality, capacity and price data from the IMF.

Note: Figure 4 includes average growth for 74 countries from 1992 to 2012 for lights and GDP. Panel A plots the bivariate correlation of the log growth of GDP and lights. Panel B conditions this relationship on the vegetation index, quality, capacity, price measurement practices, agricultural value-added in national accounts, and the share of GDP attributed to natural resources.

A7 in the online Appendix). This suggests that when the aforementioned challenges of measuring GDP are accounted for, the correlation between night-lights data and GDP is high. In short, the measurement challenges reviewed in this paper matter substantially and can help explain discrepancies in growth measurement..

How Can We Do Better?

What are some concrete steps that could improve growth measurement in developing countries? While some constraints may be political, such as policy-makers who may not be interested in statistical practices that could make them look bad, good measurement can also shine a light on progress and reveal fruitful areas for policy action. Duly noting the political constraints, we now discuss a few areas for improvement.

Improve Statistical Capacity

Improving statistical capacity is an obvious and frequent recommendation, but also a challenging one. International efforts to support national statistics offices are often focused on one-off data collection activities with limited attention to building the skills and knowledge of national statisticians or to developing data systems. Collecting data is a relatively well-defined task with a clear end date that usually wraps up with a completion report. Investments to improve statistical capacity are much more difficult to monitor, less certain to succeed, time-consuming, and often lacking clear outcome deliverables.

Infrequent GDP rebasing is one specific problem facing many developing countries that would be feasible to address. Moreover, when countries do update their GDP base years, they often do not adequately explain or document the changes; the resulting GDP volatility contributes to perceptions of possible data manipulation. While the 1993 SNA guidelines state a preference for moving away from fixed base-year methodologies towards annual chain indices, they recognize that some countries with limited statistical capacities will need to continue following fixed-base year methods. For these countries, the base year should be updated annually and then estimates should be linked across base years to maintain comparability of trend data (IMF 1993). This approach keeps reference prices (and thereby implicit weights) current, while also smoothing out discrete GDP breaks.

Combine Traditional Data with Innovative Data Sources

An explosion of new and publicly available data sources has taken place over the last decade: web-scraping, Google searches, digital transactions, mobile phone metadata, social media usage, satellite data, and others. There are important examples of these sort of data outperforming traditional data sources: for example, Blumenstock, Cadamuro, and On (2015) use mobile phone metadata to estimate poverty and wealth, and Cavallo and Rigobon (in this journal, 2016) use web-scraped price data to estimate inflation. These new sources of data are illuminating and useful

but should be viewed as complements rather than substitutes for traditional data for several reasons.

First, national income accounting relies on a wide array of data sources including data collected by other government agencies for administrative purposes, national surveys, and censuses. Most of these data were collected for purposes other than national income accounting. For example, population census, agricultural census, industrial census, price surveys, household surveys, and labor force surveys were designed for other purposes (like reducing the harms of poverty, food insecurity, and unemployment). Even if replacing a traditional data source with a new one proved successful for the narrow purpose of estimating GDP, dropping or neglecting the traditional source would most likely damage the ability to fulfill its primary purpose.

Second, traditional data sources typically seek complete coverage of current populations, although they certainly face challenges in doing so, such as underrepresentation of informal settlements, slum inhabitants, and top-income earners. In contrast, while data from new sources can be massive in sample size and very timely, they are rarely representative of the population of a nation (for example, Blank and Lutz 2017).

Third, the joint use of traditional and newer data offers complementarities, as in the examples we include in this paper of supplementing GDP measurement with satellite-based data on night lights or vegetation yields. Another example is agricultural yield measurement: while traditional fieldwork is useful for obtaining estimates of average yield, satellite data can help improve estimates of yield variations (Lobell et al. 2020). Likewise, using satellite data to augment traditional sampling frames (Tollefson 2017) based on population censuses is another example of a useful hybrid approach. The modal frequency for population censuses is once every ten years; a common practice is to survey samples at annual or more frequent intervals within geographic areas, then use the decadal census-based population weights to extrapolate annual results for the country. Sampling frames based on population censuses are often inaccurate even when fresh, because of coverage problems (particularly in densely populated areas and informal settlements), and they become outdated over time. Cross-country analysis by the Bongaarts and Bulatao (2000) finds that population counts from censuses are off by 3 percent on average in the year the count was carried out, and that the five-year projections from the base year are off by 6 percent. Supplementing population frames with a combination of satellite-based estimates of housing structures and on-ground sampling of inhabitants per typical structure can provide more accurate estimates. More accurate population estimates would have a direct role on GDP per capita estimates and could also feed into future GDP measurements.

Monitor Performance, Identify Gaps, Offer Transparency

Just as countries collect data to monitor the performance of their policies and programs, collecting metadata on national data and statistical systems would also have value. As noted, in 2021 the World Bank released the Statistical Performance

Indicators (SPI) as an upgrade to the earlier Statistical Capacity Index. Although the goal to measure the capacity of national statistical systems is the same, the new SPI has expanded into new areas including data use, administrative data, geospatial data, data services, and data infrastructure. Continuing efforts to improve the quality of assessments of data systems can identify weak links and thereby target resources for improved measurement.

In addition, the IMF regularly collects detailed information from countries on their practices with regard to the System of National Accounts, including GDP revision policies, data access prior to public release, and GDP compilation and public release practices. Much of this data exists in open-response text form and is publicly available on the IMF website for over 140 countries. As noted, the IMF recently codified a subset of this information into easy-to-analyze datasets (Berry et al. 2018; Berry et al. 2019). The IMF also conducts SNA audits, with detailed reports available publicly online for 83 countries. In this paper, we collaborated with the IMF to codify information available in these audits to create a usable dataset for the first time. Efforts similar to this one, which harness the global reach and infrastructure of institutions such as the IMF and the World Bank, could substantially improve information on national data and statistical systems.

Our analysis, and others like it, clearly show that many countries are not following the latest guidelines and compliance is far from complete. Poor transparency, including lack of commitment to open and easily-accessible data, is just as critical to address. Making data available to the public requires investing staff time and skill for documentation (including codebooks, field manuals describing protocols, sample design, and metadata on coverage and response), de-identifying and preparing the data for safe dissemination, and other steps. This requires a culture of documenting and publicly disclosing the decisions made and methodologies used in GDP estimation. Just as “sunlight is the best disinfectant,” transparency limits both the scope for and perception of political manipulation of data.

■ *We are grateful to Michael Stanger and Jim Tebrake of the IMF and to Hogeun Park of the World Bank for expert guidance and a compilation of some of the data used in this study as well as for many helpful discussions. We also thank JEP co-editors Heidi Williams and Gordon Hanson as well as Timothy Taylor for their guidance, and Josh Blumenstock, Paul Collier, Angus Deaton, Stefan Dercon, John Gibson, and David Weil for useful inputs and conversations. Rhea Gupta provided excellent research assistance and Greg Larson provided invaluable editorial support. We are grateful to the Economic Growth Center (EGC) at Yale for financial support. The views expressed here are those of the authors and should not be attributed to the World Bank.*

References

- Abay, Kibrom A., Gashaw T. Abate, Christopher B. Barrett, and Tanguy Bernard. 2019. "Correlated Non-Classical Measurement Errors, 'Second Best' Policy Inference, and the Inverse Size-Productivity Relationship in Agriculture." *Journal of Development Economics* 139: 171–84.
- Aruoba, S. Boragan, Francis X. Diebold, Jeremy Nalewaik, Frank Schorfheide, and Dongho Song. 2016. "Improving GDP Measurement: A Measurement-Error Perspective." *Journal of Econometrics* 191 (2): 384–97.
- Berry, Francien, Brian Graf, Michael Stanger, and Mari Ylä-Jarkko. 2019. "Price Statistics Compilation in 196 Economies: The Relevance for Policy Analysis." International Monetary Fund Working Paper 19/163.
- Berry, Francien, Massimiliano Iommi, Michael Stanger, and Louis Venter. 2018. "The Status of GDP Compilation Practices in 189 Economies and the Relevance for Policy Analysis." International Monetary Fund Working Paper 18/37.
- Blank, Grant, and Christoph Lutz. 2017. "Representativeness of Social Media in Great Britain: Investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram." *American Behavioral Scientist* 61 (7): 741–56.
- Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting Poverty and Wealth from Mobile phone Metadata." *Science* 350 (6264): 1073–76.
- Bongaarts, John, and Rodolfo A. Bulatao, eds. 2000. *Beyond Six Billion: Forecasting the World's Population*. Washington, DC: National Academy Press.
- Buddenberg, Doris, and William A. Byrd, eds. 2006. *Afghanistan's Drug Industry: Structure, Functioning, Dynamics, and Implications for Counter-Narcotics Policy*. Washington, DC: United Nations Office on Drugs and Crime.
- Capéau, Bart, and Stefan Dercon. 2006. "Prices, Unit Values and Local Measurement Units in Rural Surveys: An Econometric Approach with an Application to Poverty Measurement in Ethiopia." *Journal of African Economies* 15 (2): 181–211.
- Carletto, Calogero, Dean Jolliffe, and Raka Banerjee. 2015. "From Tragedy to Renaissance: Improving Agricultural Data for Better Policies." *The Journal of Development Studies* 51 (2): 133–48.
- Carletto, Calogero, Sara Savastano, and Alberto Zezza. 2013. "Fact or Artifact: The Impact of Measurement Errors on the Farm Size–Productivity Relationship." *Journal of Development Economics* 103: 254–61.
- Carson, Carol S., Charles Enoch, and Claudia Helene Dziobek, eds. 2002. *Statistical Implications of Inflation Targeting: Getting the Right Numbers and Getting the Numbers Right*. Washington, DC: International Monetary Fund.
- Cavallo, Alberto, and Roberto Rigobon. 2016. "The Billion Prices Project: Using Online Prices for Measurement and Research." *Journal of Economic Perspectives* 30 (2): 151–78.
- Charmes, Jacques. 2012. "The Informal Economy Worldwide: Trends and Characteristics." *Margin: The Journal of Applied Economic Research* 6 (2): 103–32.
- Chen, Wei, Xilu Chen, Chang-Tai Hsieh, and Zheng Song. 2019. "A Forensic Examination of China's National Accounts." NBER Working Paper 25754.
- Chen, Xi, and William D. Nordhaus. 2011. "Using Luminosity Data as a Proxy for Economic Statistics." *Proceedings of the National Academy of Sciences* 108 (21): 8589–94.
- Chodorow-Reich, Gabriel, Gita Gopinath, Prachi Mishra, and Abhinav Narayanan. 2018. "Cash and the Economy: Evidence from India's Demonetization." NBER Working Paper 25370.
- Deaton, Angus. 2005. "Measuring Poverty in a Growing World (or Measuring Growth in a Poor World)." *Review of Economics and Statistics* 87 (2): 395.
- Deaton, Angus, and Salman Zaidi. 2002. "Guidelines for Constructing Consumption Aggregates for Welfare Analysis." LSMS Working Paper 135.
- Desiere, Sam, and Dean Jolliffe. 2018. "Land Productivity and Plot Size: Is Measurement Error Driving the Inverse Relationship?" *Journal of Development Economics* 130: 84–98.
- Devarajan, Shantayanan. 2013. "Africa's Statistical Tragedy." *Review of Income and Wealth* 59 (S): S9–15.
- Dillon, Andrew, Sydney Gourlay, Kevin McGee, and Gbemisola Oseni. 2018. "Land Measurement Bias and Its Empirical Implications: Evidence from a Validation Exercise." *Economic Development and Cultural Change* 67 (3): 595–624.
- Donaldson, Dave, and Adam Storeygard. 2016. "The View from Above: Applications of Satellite Data in

- Economics." *Journal of Economic Perspectives* 30 (4): 171–98.
- The Economist.** 2014. "How Nigeria's Economy Grew by 89% Overnight," *The Economist*, April 8. <https://www.economist.com/the-economist-explains/2014/04/07/how-nigerias-economy-grew-by-89-overnight>.
- Elvidge, Christopher D., Edward H. Erwin, Kimberly E. Baugh, Daniel Ziskin, Benjamin T. Tuttle, Tilottama Ghosh, and Paul C. Sutton.** 2009. "Overview of DMSP nighttime lights and future possibilities." In *2009 Joint Urban Remote Sensing Event*, 1–5. Piscataway, NJ: IEEE.
- Gibson, John, Susan Olivia, Geua Boe-Gibson, and Chao Li.** 2021. "Which Night Lights Data Should We Use in Economics, and Where?" *Journal of Development Economics* 149.
- Gourlay, Sydney, Talip Kilic, and David B. Lobell.** 2019. "A New Spin on an Old Debate: Errors in Farmer-Reported Production and Their Implications for Inverse Scale–Productivity Relationship in Uganda." *Journal of Development Economics* 141.
- Henderson, J. Vernon, Adam Storeygard, and David N Weil.** 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028.
- International Labor Organization.** 2004–2016. "Proportion of Informal Employment in Total Employment by Sex and Sector (%)." *International Labor Organization* https://www.ilo.org/shinyapps/bulkexplorer25/?lang=en&segment=indicator&id=SDG_0831_SEX_ECO_RT_A. (accessed September 7, 2020).
- International Labor Organization.** 2018. *Women and Men in the Informal Economy: A Statistical Picture*. Geneva: International Labour Office.
- International Monetary Fund.** 1993. *System of National Accounts, 1993*. Washington, DC: International Monetary Fund.
- International Monetary Fund.** 2001–2019. "Reports on the Observance of Standards and Codes (ROSCs)." <https://www.imf.org/en/Publications/rosc>.
- Jerven, Morten, and Deborah Johnston.** 2015. "Statistical Tragedy in Africa? Evaluating the Data Base for African Economic Development." *The Journal of Development Studies* 51 (2): 111–15.
- Kerner, Andrew, Morten Jerven, and Alison Beatty.** 2017. "Does It Pay to Be Poor? Testing for Systematically Underreported GNI Estimates." *Review of International Organizations* 12: 1–38.
- Koren, Miklós, and Silvana Tenreyro.** 2007. "Volatility and Development." *Quarterly Journal of Economics* 122 (1): 243–87.
- Kuznets, Simon.** 1934. *National Income, 1929–1932*. Cambridge, MA: National Bureau of Economic Research.
- Lobell, David B., George Azzari, Marshall Burke, Sydney Gourlay, Zhenong Jin, Talip Kilic, and Siobhan Murray.** 2020. "Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis." *American Journal of Agricultural Economics* 102 (1): 202–19.
- Lowder, Sarah K., Jakob Skoet, and Terri Raney.** 2016. "The Number, Size, and Distribution of Farms, Smallholder Farms, and Family Farms Worldwide." *World Development* 87: 16–29.
- Medina, Leandro, and Friedrich Schneider.** 2018. "Shadow Economies around the World: What Did We Learn Over the Last 20 Years?" IMF Working Paper 18/17.
- National Centers for Environmental Information.** 2012. "National Trends in Satellite Observed Lighting: 1992–2012". *National Oceanic and Atmospheric Administration* https://ngdc.noaa.gov/eog/dmsp/download_national_trend.html (accessed September 7, 2020).
- Nussbaum, Martha.** 1987. "Nature, Function, and Capability: Aristotle on Political Distribution." WIDER Working Paper 31.
- Open Street Map.** 2017. "ne_50m_admin_0_countries." https://puma.worldbank.org/layers/geonode%3Ane_50m_admin_0_countries (accessed September 7, 2020).
- Pinkovskiy, Maxim, and Xavier Sala-i-Martin.** 2016. "Lights, Camera...Income! Illuminating the National Accounts-Household Surveys Debate." *The Quarterly Journal of Economics* 131 (2): 579–631.
- Prydz, Espen, Jolliffe, Dean and Umar Serajuddin.** 2020. "Mind the Gap: Disparities in Assessments of Living Standards Using National Accounts and Surveys." Conference paper, IARIW-WB Conference on New Approaches to Defining and Measuring Poverty in a Growing World, Washington, DC.
- Ravallion, Martin.** 2003. "Measuring Aggregate Welfare in Developing Countries: How Well Do National Accounts and Surveys Agree?" *Review of Economics and Statistics* 85 (3): 645–52.
- Sen, Amartya.** 1985. *Commodities and Capabilities*. Amsterdam: North-Holland.
- Steindel, Charles.** 1995. "Chain-Weighting: The New Approach to Measuring GDP." *Current Issues in Economics and Finance* 1 (9): 1–6.

- Stiglitz, Joseph, Amartya Sen, and Jean Fitoussi.** 2009. *Report of the Commission on the Measurement of Economic Performance and Social Progress (CMEPSP)*. <https://ec.europa.eu/eurostat/documents/8131721/8131772/Stiglitz-Sen-Fitoussi-Commission-report.pdf>.
- Stone, Richard.** 1947. "Measurement of National Income and the Construction of Social Accounts." Report of the Subcommittee on National Income Statistics of the League of Nations Committee of Statistical Experts 7. Studies and Reports on Statistical Methods. Geneva, Switzerland: United Nations.
- Stone, Richard, D. G. Champernowne, and J. E. Meade.** 1942. "The Precision of National Income Estimates." *Review of Economic Studies* 9 (2): 111–25.
- Subramanian, Arvind.** 2019. "India's GDP Mis-Estimation: Likelihood, Magnitudes, Mechanisms, and Implications." CID Faculty Working Paper 354.
- Tollefson, Jeff.** 2017. "Satellite Images Reveal Gaps in Global Population Data." *Nature* 545 (7653): 141–42.
- US Geological Survey.** 2001–2018. "Landsat Normalized Difference Vegetation Index." https://www.usgs.gov/core-science-systems/nli/landsat/landsat-normalized-difference-vegetation-index?qt-science_support_page_related_con=0#qt-science_support_page_related_con (accessed September 7, 2020).
- Wilson, Tom, and David Blood.** 2019. "Rwanda: Where even Poverty Data Must Toe Kagame's Line," *Financial Times*, August 12. <https://www.ft.com/content/683047ac-b857-11e9-96bd-8e884d3ea203>.
- World Bank.** 1960–2019. "Agriculture, Forestry, and Fishing, Value Added." <https://data.worldbank.org/indicator/NV.AGR.TOTL.ZS> (accessed September 7, 2020).
- World Bank.** 1960–2019. "World Development Indicators." <https://data.worldbank.org/indicator/NY.GNP.PCAP.CD> (accessed September 7, 2020).
- World Bank.** 1960–2019. "Poverty & Equity Data Portal." *PovcalNet* <https://povertydata.worldbank.org/poverty/home/> (accessed September 7, 2020).
- World Bank.** 2004–2020. "Statistical Capacity Indicators." <https://databank.worldbank.org/reports.aspx?source=Statistical-capacity-indicators> (accessed September 7, 2020).
- Young, Alwyn.** 2012. "The African Growth Miracle." *Journal of Political Economy* 120 (4): 696–739.
- Zhuo, Li, Jing Zheng, Xiaofan Zhang, Jun Li, and Lin Liu.** 2015. "An Improved Method of Night-Time Light Saturation Reduction Based on EVI." *International Journal of Remote Sensing* 36 (16): 4114–30.