# Development by Numbers – A primer

Morten Jerven[1]

How good are the numbers? This paper provides a review of the data quality in the most important databases on economic development. It discusses the provenance and quality of the observations in the data sets and equips data users with a guide to judge data quality. The adequacy of the most influential databases is assessed, and the paper furnishes guidance on how to use and how not to use the numbers. The paper explains the sources of the data, and highlights the most serious gaps in our knowledge about economic development. The paper discusses different methods of gauging data quality and whether or not we can assert a margin of error to different data and databases.

[1]Associate Professor in Global Change and International Relations, International Environment and Development Studies (Noragric), Norwegian University of Life Sciences, www.mortenjerven.com

## I. Introduction

Our knowledge of development through numbers is much more limited than their frequent use would indicate. Arguably, many of the econometric models in use for testing hypothesis are much more sophisticated than the quality of the data would justify (Ward, 1971; 2004). It is hardly a new problem. One of the pioneers in the study of economic development, Dudley Seers (1952-53, p. 160) was considerably pessimistic about the rewards of instituting national accounting for the purpose of international comparisons of income and economic development. "In the hands of authorities, such international comparisons may yield correlations which throw light on the circumstances of economic progress, and they tell us something about relative inefficiencies and standards of living, but they are very widely abused. Do they not on the whole mislead more than they instruct, causing a net reduction in human knowledge?" In the 1970s, Hirschman lamented that "the spread of mindless number-work in the social sciences has been caused largely by the availability of the computer" (1970, p. 329-43). Striking a bit more enthusiastic tone, in a 1989 survey of development economics research, Nicholas Stern described the availability of the global datasets on economic growth, such as the Penn World Tables as 'a major public good and an important statistical event' (1989, p. 600)[2]. The availability of the database was in effect expanding the boundary for empirical investigation by numbers.

Only a few years later, in 1994, T. N. Srinivasan edited a special issue of the Journal of Development Economics, which surveyed the knowledge problems presented by the use of these new statistical databases. Srinivasan raised the "concern that analyses based on unreliable and biased data could result in seriously distorted, if not altogether wrong, analytical and policy conclusions" (1994, p. 4-5). The volume contained a critical survey of some of the key statistics used to study economic development. The instigation then was the sheer increase in data on economic development. Several decades on, things have, of course, changed,[3] but the numbers are still soft, and statistical challenges remain. After a few decades of neglect, some recent well-publicized statistical events have led to an increase in the attention being paid to

[2] The first versions of the Penn World Tables were published in the 1970s (Kravis, Heston, and Summers 1978), but the mainstream use of this dataset is dated to the 5.0 version, which was published in 1991 (Summers & Heston, 1991).

[3] In 1989, Stern noted that the third and fourth editions of the World Tables (1983, 1987) 'are now available on magnetic tape' – the ease of access to data has increased dramatically,

the quality of macroeconomic statistics in LICs, especially in African countries, with the World Bank's chief economist for Africa recently declaring "Africa's Statistical Tragedy" (Deverajan, 2013; see Jerven & Johnston, 2015d).

The persistent doubts about developing countries' ability to provide valid statistics may in part be a true reflection of the data, but it may also be a perception and credibility problem. It is not as if 'rich countries' do not have problems with the credibility of their statistics. Recent stories emerging from China, Greece or Argentina, should indeed remind us that the phrase 'lies, damned lies and statistics' means that such problems are ubiquitous. We often use the term 'valid' when we speak of whether statistics are correct or not, and it is instructive that the root of the word 'valid' is similar to the root of the word 'power' (Porter, 1995). The reason we trust numbers is not only due to technical accuracy, but as much to the social power, legitimacy, and credibility of the institution that provides the facts.

The problem of manipulation of statistics, plain errors and the simple knowledge problem that not all that counts can be counted is general to the social sciences. However, I have argued that when it comes to studying economic development, our knowledge based on numbers is doubly biased: we know little about poor countries and even less about the poor people who live in these countries (Jerven, 2013a). That is a grave diagnosis. Particularly, if you think that the main purpose of organizations like the World Bank investing in, collecting and disseminating statistics is in order to obtain actionable knowledge to alleviate poverty and aid economic development. These problems emerge from a variety of sources. At the design level, an incompatibility exists between statistical categories that were conceived for industrialized societies (with clearly defined property rights and formal employment relationships) and the developing contexts to which they are applied. At the implementation level, a lack of capacity and record keeping at official statistical offices is exacerbated by the challenge of inaccessibility that is associated with poor and remote areas. Thus, numbers and indicators are especially inadequate in less developed countries.

Studying statistics is similar to studying states. Just as one would question whether one should expect the emergence of a state similar to the Japanese or Norwegian state in low income countries, one would also be careful in treating statistics from different state systems as factual equivalents. The blueprints of statistical systems are based on states that are registering and taxing land and registering and taxing citizens. In many countries land has typically not been subject to private property rights. Many states have not collected taxes on land holdings, and for most countries, vital statistics and civil registration have remained incomplete to this data. Obviously, this has direct implications for the power of the state (Herbst, 2000). The manner in which the colonial and post-colonial African state adapted to the inability to control access to land has been summarized in the concept of the 'gatekeeper state' (Cooper, 2002). Unable to collect taxes on land, income, or production, the state settled on collecting taxes at the port by levying duties on exports and imports.

This system of control and power is mirrored in the system of information. The result is that in many low income countries the data basis for aggregating measures of income and growth is weak. For large sectors of the economy we have little or no information, and the aggregate figures involve a great deal of guessing. It also means that the statistical systems are more dependent on the quality and regularity of survey data. These data are expensive. Because statistical offices suffer from low domestic funding and their budgets are dominated by donor funds,[4] it has also meant that survey ability has been relatively low and that priorities have varied with the priorities of those who pay for surveys (Jerven, 2013a; Jerven & Johnston, 2015d; Glassman & Sandefur, 2015). Moreover, sometimes the survey results, even if they are regular and well designed, may still yield misleading information (Jerven, 2015a).

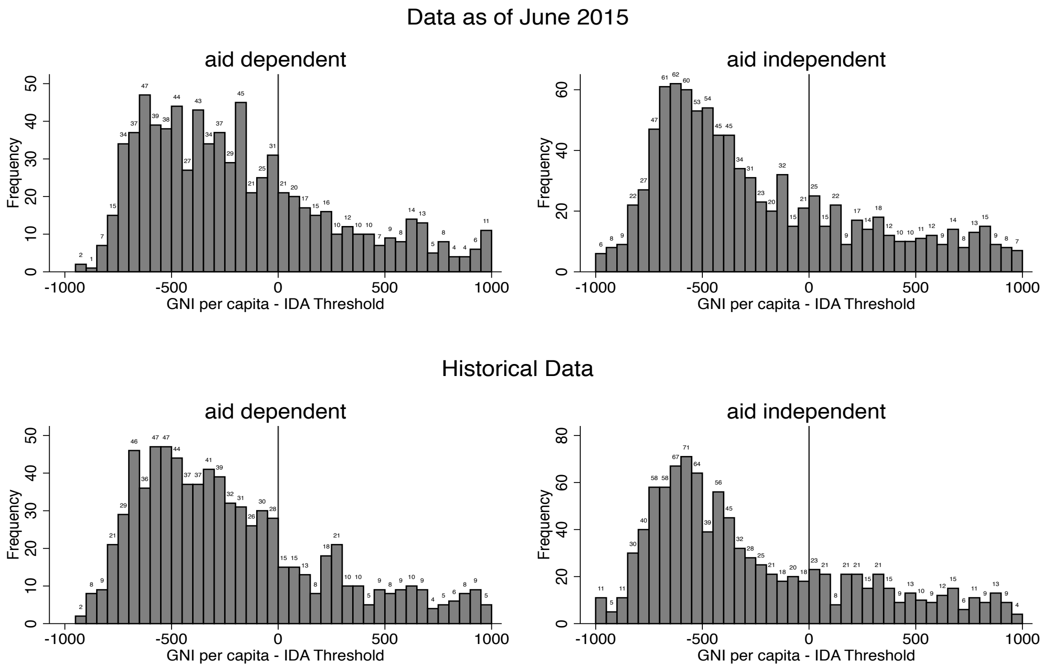
Numbers and indicators are especially inadequate in less developed countries. The paradox here is that in these very countries, numbers have a greater importance. These countries are more dependent (both politically and financially) on international organizations and global governance – an arena where numbers are key motivators of the political debate. The policy circle in international development is often designed and presented as apolitical – where funders 'pay for results' (Birdsall et al., 2010). It is dominated by technocrats, donors and international organizations that may initiate, change, or abort policies based on these feeble statistics. Indeed, the criteria which determines which countries are in the low-income group, and thus eligible for concessional lending, depends on whether countries are measured as being below or above a certain income threshold (currently 1,200 US GNI per capita). As one might

[4] Just like in the case of many other public goods provided by states, such as judicial systems and a national defense, there are economics of scale at play, so that in a relative sense it is more expensive for Burundi to have a well-functioning statistical office, compared to the small relative share it takes of the national budget in Belgium.

expect, this also has an effect of clustering countries just below the threshold over time (Kerner, Jerven & Beatty, 2015).

Initial eyeballing of the data, and then econometric tests confirm that the distribution is discontinuous around the threshold. With a simple histogram it is not hard to accurately guess the location of the IDA eligibility threshold based solely on the distribution of Atlas GNI per capita data. The figures show that observations of countries being just below the IDA threshold (-$50 and $0) are almost twice as numerous as countries just above ($0 and $50). Similarly observations between -$100 and $0 are similarly nearly twice as numerous as between $0 and $100. One would not expect to see the same clear discontinuity in the WDI data (because it does not determine aid directly) as its effect is less pronounced. Moreover, one would expect to see it in the GNI per capita data, but only for the countries that are aid dependent countries.

Figure 3: Distribution of GNIs per capita around IDA threshold 1987-2013, by aid dependence




Beyond the policy domain, the deeper methodological point for scholars and researchers is that numbers need to be interrogated meticulously. Confronted with secondary data in the international databases, users need to conduct basic source criticism and ask ‘who made this observation?’, ‘under what conditions was this observation made?’, and ‘is there any reason to think that the observation is biased?’

Failure to do so increases the distance between the observer and the observed and may lead to a disconnect between reality and the numbers.

It is a striking feature of the social sciences that we would attach little value to an international businessman stating that 'Tanzania is more corrupt than Zambia'. However, if Transparency International or the Economist report that the numerical average of such vague, arbitrary and subjective statements resulted in Tanzania and Zambia respectively scoring 4,2 and 3,9 on a scale from 1 to 6 where 6 is most corrupt, most data users would be happy to use this observation as an established social fact.

In this paper I first discuss how and why poor numbers matter. I then offer a general discussion of how statistics are made, and trace how numbers travel from the field surveys to statistical offices, and through to the final dissemination by international datasets. I then provide a guide to the most important metrics used in studying development. I summarize by providing an assessment of the different datasets. It is clear that the numbers are soft, just how soft would be of interest to researchers, whose conclusions will depend on the firmness of these facts of economic development.

## II. How poor numbers harm research and policy

Poor numbers may harm policy research and policy decisions. It affects research in three obvious ways. First, through false negatives, second through false positives, and finally through missing data. An example of a false negatives is that there may be a significant relationship between corruption and economic growth, but because corruption and economic growth are both poorly measured, the relationship is undetectable on average using cross sectional correlations, and thus research may be communicated as 'we have not found a relationship' which in turn is often interpreted as 'there is no relationship'. These errors are endemic in social science research – but may be particularly prone to show up if you run regressions with bad data on both sides of the equation. The existence of such false negatives or not have been subject to a lot of debate in the literature that looks for a relationship between aid and growth (as discussed in Easterly, 2003)

False positives are often believed to be less common. That is, we do accept that pregnancies can go undetected, but we do not want our tests to signal pregnancies in cases where the subject is not pregnant. The standard defense is simply that if errors are random, then it would bias the results towards false negatives. Thus, if you have a positive, it is unlikely that it is driven by bad data. So there is at least some comfort in that data errors are in part biased against finding a relationship.[56] Unfortunately, that is not always true, particularly if we are looking at data that are subjective (such as reported answers to questions like: 'on a scale from 1 to 10 how corrupt do you think Nigeria is?'), where we would think that the source of the data and the researcher might have similar preconceptions. Because, in most cases, there is no check on the divergence between 'subjective' and 'objective' data the importance of such false positives is difficult to assess. An example would be to compare the 'perception of corruption' with actual observed cases of corruption. But since 'revealed cases of corruption' would not be a measure of 'all cases of corruption', and since it is likely that the two variables 'perception of corruption' and 'revealed cases of corruption' would have a causal relationship with each other, this remains a difficult hypothesis to test. It is indicative of such problems that perceptions of corruption are almost perfectly correlated with levels

[5] Though that bias is surely overcome by the practice of scholars of re-running specifications until a relationship is actually found, despite problems of the data (Leamer, 1983; Easterly, 2009).

of income (Cobham, 2014) and that the relationship between economic performance and corruption is weaker when you use 'revealed cases of corruption' rather than 'perception of corruption' (Aidt, 2010).

Data are also unlikely to be random in the social sciences because the numbers are created through social processes. It is said that as soon as an indicator becomes important it also becomes useless. The subtle effects are often referred to as the 'Hawthorne Effect' or the observer effect. It describes the phenomena that individuals modify or improve an aspect of their behavior in response to their awareness of being observed. Indeed, a lot of statistical collection is implicitly motivated by this effect. One of the reasons that the development community is propagating new metrics, new rankings and new targets such as the Human Development Index, the poverty headcount, the Millennium Development Goals, and the Sustainable Development Goals is not monitoring for monitoring's sake, it is intended to change behavior.

Of course, the hope is that actors respond to the metric by improving their behavior, but that assumes a very clean measurement process. It is often easier to imagine that behavior may be directed towards the measurement process itself. In other words, as any teaching and grading research professor would know, it is likely that you will have behavior that enhances the measurement, but which is not necessarily reflected in actual performance. Campbell's Law, the not so distant and more harmful relative of the Hawthorne Effect, dictates that: "The more any quantitative social indicator (or even some qualitative indicator) is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor." (Campbell, 1976)

Thus, there are obvious reasons why one should be careful about false positives as well. Did a donor aided project targeting primary school enrollments succeed in increasing primary school enrollment? The researcher is asserting a high level of objectivity to the numbers of school enrollments coming out of the administrative statistical systems if those are trusted, particularly if incentives are clearly stated at the beginning of the program.

Sandefur and Glassman (2015) study an example from education data in Kenya. A donor supported intervention abolished school fees and financially rewarded Kenyan authorities for putting students through primary school. According to administrative data there was a rapid increase in enrolment following the abolishment of fees. The evidence showed the policies worked. But donor-funded surveys showed a different picture of school enrolments. The Demographic and Health Survey (DHS) showed enrolment rates that were flat over time. The discrepancy owed to the fact that schools got more funding if they reported more pupils. Significant here is that the discrepancy between the administrative data and the actual enrolment rates per the DHS emerged after the intervention that paid for results. In sum, the policy intervention had only one effect that could be clearly established: as a result of the policy the Kenyan government had no reliable evidence on how many children actually went to school.

Before dealing with missing data, which has pretty obvious effects, it is worth considering phenomena that are not well quantified, such as 'governance,' or 'rights,' or even more mundane 'poverty,' or simply 'development'. Most of the time scholars end up using more or less explicit proxies. It is a truism of statistics that, when using a proxy in place of a real phenomenon, the former will eventually substitute for the latter. From this, the policy will target an indicator that stands in for the phenomenon that is actually being targeted (Cooley & Snyder, 2015). This can happen in one of two ways; either as a narrowing of policy that ignores other important issues (like the causal mechanisms that lead to poverty), or worse, as an indicator that incentivizes agents to act to influence the indicator rather than the problem.

In a review of the effect of quantification in development, Fukada-Parr and Yamin (2013) go through the most prominent MDGs and point out many such unintended consequences of the MDGs – in particular they emphasize the reductionist tendency of quantification. So that for example, Goal 1, Target B aimed to halve the proportion of the population that is undernourished. Undernourishment drove a caloric consumption dimension of hunger, and was reinforced by the second indicator of weight. These metrics neglect other dimensions of food insecurity including under-nutrition and insecurity. Alternative indicators, such as heights or price volatility in national price indices, would have highlighted dimensions which approach food insecurity as a longer term challenge, and calls for a different policy approach.

All these indicators are at some level entering the realm of 'as if'. On some level all data users know that the Freedom House actually does not measure 'democracy'; that the Consumer Price Index does not actually measure 'inflation'; nor does Transparency International actually measure 'corruption'. We just pretend 'as if' they do. If researchers or policy makers forget that we are making decisions or doing analytic research with the important caveat that we are acting 'as if' these things can be counted, it will lead us astray. Abdelal and Blyth (2015) make the case, general to subjective global rankings, that the Credit Rating Agencies fail to act as a judge – making an objective assessment of a balance sheet. The balance sheets can be unchanged, yet rankings can change, not reflecting financial solidity as such, but responding to, and thereby triggering subjective notions of risk. A particular case of subjective metrics becoming 'objective' through legislative action was reported on by Matt Collins (2015). The UK's Financial Conduct Authority fined the British branch of the Bank of Beirut for lack of sufficient controls against money laundering, and barred it from accepting more customers in "high-risk" jurisdictions. 'High risk' was defined as any country scoring 60 or below on Transparency International's Corruption Perception Index. Of the 175 countries that are ranked by Transparency International, nearly 80 percent are below this threshold. However, Transparency International is not monitoring money laundry, they are collecting subjective impressions of public sector corruption.

Surely, this latter general point is extendable to all aspects of social inquiry, but as scholars of economic development, one should perhaps be even more mindful of this, particularly in studies which are not grounded in fieldwork, but are based on borrowed data and numbers from downloaded datasets. It is perfectly conceivable, for instance, to write a paper on the economic effects of civil wars in Africa without having set foot in any of the countries, by simply relying on data from the World Bank and the Uppsala Conflict Data Program. Arguably, with the combined availability of powerful computers and downloadable datasets, the distance between the observer and the observed continues to increase since the 1990s. The monographic field-based study of the single economy gave way to article-focused inquiry taking advantage of cross-country sectional variation to tease out why some countries are performing well and others remain poor (Jerven, 2015b).

## III. Trends in use of numbers in development

In 2013, the UN High Level Panel (2013) delivered its report with recommendations for Sustainable Development Goals subsequently to be adopted by the UN General Assembly in 2015. One small aspect of the report very soon caught everyone's attention. Buried on page 8 in the report there was a call for a 'Data Revolution' in development. The promise of measurability generated a frenzy of enthusiasm in the international development community. One year later, the Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (2014) put forward its recommendation with the title *A World That Counts*.

The report laid out a grand ambition. It recognized that currently "whole groups of people are not being counted and important aspects of people's lives and environmental conditions are still not measured." From that acknowledgement came a startling next step with the report declaring that, "[n]ever again should it be possible to say, 'we didn't know'. No one should be invisible. This is the world we want – a world that counts." One can understand this enthusiasm, though with some reservation. There are serious limits to what can be known through counting. Moreover; not all issues can be resolved through counting. Restricting the production of knowledge and the design of governance to these methods of numeracy and counting may have serious pitfalls. Counting is not the same as knowing and, though this might be implicitly acknowledged, the practical needs of policymakers and researchers may tempt us to overemphasize the quantifiable. The reality is that you need more discretionary and pragmatic approaches that take into account other sources of actionable knowledge.

Overwhelmingly, the dominant strategy in the international development community has been to forge ahead and increase our reliance on numbers. Rather than abiding by the need to exercise caution when using numbers; more numbers, indicators, and targets are demanded. This is not surprising in and of itself - we trust numbers because they appear objective, factual, and easily interpretable – they are presented as 'hard facts' rather than soft, subjective, anecdotal, and qualitative observations.

In conclusion, the demand for and dependence on statistics and indicators is on the rise in the development community. The aims of development are increasingly stated in quantifiable metrics – first with the Millennium Development Goals (Black and

White, 2004) and now with a more ambitious agenda in the Sustainable Development Goals. The buzz in the development community is around 'evidence based policy'. This impression of measurability and accuracy is misleading. Instead, research has documented a tendency towards the inverse – 'policy based evidence' (Jerven, 2014b). This is the case when a particular policy objective endogenously creates evidence that is used in place of an exogenous examination of whether the policy worked or not (Boden & Epstein, 2006).

For example, when donors set a target to increase primary school enrollment with a cash reward, administrative structures might respond by reporting higher enrollment in order to qualify for those very rewards (as has already been documented in education Kenya by Sandefur and Glassman (2015), who report similar findings on vaccinations in Tanzania. In the case of Malawi, the government introduced fertilizer subsidies in form of a voucher. The program was donor funded, and the government was under pressure to show that yields were increasing. The success was celebrated in the New York Times by Jeffrey Sachs (2009). The widely circulated figures highlighted that the maize harvest in 2006/2007 was as high as 3.4 million tonnes. An agricultural census conducted in the same year (2006/2007), the publication of which was much delayed by the authorities, later indicated a maize output of 2.1 million tonnes, only 60 percent of what was previously reported by the Ministry of Agriculture. The discrepancy of 1.3 million tonnes was driven by a much higher number of farms used as the multiplier in the Ministry of Agriculture. There were shared incentives at the top and at the bottom. The President and the Ministry desired good, consistent performance in order to keep the electorate and donors convinced of the continued success of the agricultural development strategy. It was also in the interest of small-holders and agricultural extension officers to 'increase' the numbers of farming households, not only to please superiors, but also because the vouchers themselves have a market value (Jerven, 2014f).

Numbers do not appear from thin air; they are embedded in social and political discourse, and their provision has a cost (Jerven, 2014c). Such an acknowledgment has broad implications across the social sciences and particularly in the context of developing countries. There is a fundamental failure to understand that the statistics – the so-called objective facts – are social constructs and political products and that we need to understand the political economy that surrounds them.

More fundamentally, counting is not synonymous with knowing. The act of counting does not guarantee objectivity nor does it inherently make us wiser. In the great majority of cases, counting, or the process of quantification, involves some standardization and categorization. While we gain knowledge from aggregation and addition, we also lose information in the same process. One oft-discussed example of this is the world poverty headcount. We devote a large amount of resources to calculating this global number – how many people live below $1.25 a day – but it is a curious creation (Subramanian, 2012). First, the number is a global aggregate but is based on very small and sometimes entirely non-existent local samples. As is documented below, and elsewhere (Serajuddin et al, 2015), household surveys covering a few thousand households are only conducted in some countries some of the time. Projections across time and space are then aggregated in a way that precludes the calculation of poverty statistics at anything but the global level. Thus, as an indicator, you can react to it and you can base an advocacy campaign or a media story on it, but you cannot design policy around it. This is because the information in the indicator does not contain any information about what causes a change in the phenomena that it is reporting on. This dearth of information is not an inherent, general, or unavoidable property of such a statistic. On the contrary, the budgets created in financial accounting give you an aggregate, which can tell you whether you are running a deficit or not. Further, it can be easily disaggregated so that you can determine exactly what has changed to cause a deficit in a particular accounting year. Thus, there are different indicators that can be used for different purposes.

Little thought has been put into asking why we should invest in statistics and, if so, what kind. The UN report says demanding more data will lead to better decisions. This is a normative statement and ignores a conceptualization of how evidence relates to policy. The causal relationship between more statistics and better decisions is not clarified. The growth of statistical systems in the 20th century suggests that the body of statistics that we have today is a result of policy decisions made previously. Meaning that statistics often turn out to be an unintended or strategically generated fingerprint of the state's activities. If a state has no agricultural policy, it is likely that it has no agricultural statistics to speak of. If the states are not able to tax the unrecorded or informal sector, it remains unrecorded. Thus, the statistical record of a country mirrors the political priorities (Jerven, 2011c). Empirical and theoretical research need to investigate if influencing a state's manner of collecting and disseminating information will change the way that it behaves. We do not actually know how a change in a statistics will change policy. This uncertainty casts doubt over how much we invest in these numbers and what type of metrics we should be investing in.

How important are these statistics for scholarly work? As mentioned, the real moment came in the 1990s, when the combination of a new methodology, better technology and new data sources spurred a great amount of research. Particularly notable, within economics and political science, was research using cross-country growth regressions in which the dependent variable was the average growth rate of per capita GDP. Researchers added different independent variables, or interactions of independent variables, to the initial baseline estimation in their search for new insights about what correlates with economic growth (Durlauf, Johnson, & Temple, 2005). Durlauf and colleagues referred to this scholarly production over the following decade as a 'growth regression industry' (Durlauf, Johnson, & Temple, 2005, p. 599),[7] and this was in large part made possible by statistical events such as the Penn World Tables becoming available.

In mainstream economics there has been a radical shift from theory-based to data-based research publishing. In a survey of leading economic journals (American Economic Review, Journal of Political Economy, and Quarterly Journal of Economics), Hamermesh (2013) highlighted a shift away from theoretical articles and towards data-use. Beginning in the 1960s, the author took a sample year from each decade and analyzed the full-length refereed articles published in the three journals that year. This led to a sample of 748 articles collected from publications in 1963, 1973, 1983, 1993, 2003, and 2011. Hamermesh's results are reproduced below. Evidenced is a decline in theoretical publications in place of empirical ones. While empirical publications with borrowed data (data copied from books, or provided electronically) peaked in the 1990s, empirical articles using self-generated datasets has continued to rise recently. The rise of experiments and use of own data may indeed be seen as a response to the increasing uncertainty attached to international and national data sources, or 'borrowed data' as Hamermesh calls it.

[7] According to a review of the growth literature to date, 145 explanatory variables have been found to be statistically significant (Durlauf, Johnson, and Temple, 2005, Appendix 2) and therefore can help explain the rate of growth. Of these 145 variables, some entertain similar growth hypotheses but differ in the measures used. The authors identified 43 conceptually different theories about economic growth as being 'proven' in the literature (ibid., p. 639).

**Table 1**

| Percent Distribution of Methodology of Published Articles, 1963-2011* | | | | | |
|---|---|---|---|---|---|
| | Type of Study | | | | |
| Year | Theory | Theory w/ Simulation | Empirical: Borrowed Data | Empirical: Own Data | Experiment |
| 1963 | 50.7 | 1.5 | 39.1 | 8.7 | 0 |
| 1973 | 54.6 | 4.2 | 37.0 | 4.2 | 0 |
| 1983 | 57.6 | 4.0 | 35.2 | 2.4 | 0.8 |
| 1993 | 32.4 | 7.3 | 47.8 | 8.8 | 3.7 |
| 2003 | 28.9 | 11.1 | 38.5 | 17.8 | 3.7 |
| 2011 | 19.1 | 8.8 | 29.9 | 34.0 | 8.2 |
| *A type could not be assigned to seventeen of the articles published in 1963. | | | | | |

Source: Hamermesh (2013, p. 168).

Statistical indicators are more influential than ever before. The Economist provided one recent overview of the power of indicators in a report on 'How to lie with indices' (2014). There are more numbers that are having more influence, and so we need to get a handle on how the numbers are produced and what kind of power they have. There is a rise in indicators. Moreover, fueled by Big Data and the Sustainable Development Goals, the so-called 'data revolution' is finding its feet, so one should not expect it to let up anytime soon.

As one would expect there is now a ballooning literature to match the rise of indicators. The early and leading contributions in the literature on indicators emerged from Law and Anthropology (sometimes cross-disciplinary work), but now Political Science, and in particular International Relations, is following. Economics is, as far as I can tell, still on the sidelines. Of course, 'Constructivism' within International Relations has a natural comparative advantage in approaching data as social products. On the other hand, economists, statisticians, and political scientists of the positivist mold, have a mountain to climb. Despite the many publications on this topic, I think there are still many holes in our knowledge. There is a further need for empirical research on the lines of political ethnography of indicators. Particularly there is a knowledge gap in the study of the relationship between data and decisions. There is also a surprising gap in knowledge of what makes a good indicator and what does not.[8]

It would be naïve to think that the use of numbers in research, however misleading, will not continue. The underlying issue here is that numbers need to be interrogated meticulously. In 2016 we are at a point where we have the legacy of two decades of such cross country database work. Looking backward and forward, we are also in the time of the 'data-revolution' in which the importance of evidence, and particularly numerical evidence drawing from official or international organizations as sources, will likely increase in importance. Thus, in the following, I will evaluate the availability, reliability, and conceptual soundness of the numbers that provide the backbone of the study of development by numbers. But first we need to understand where the data come from.

[8] Broome and Quirk has collected a set of papers evaluating what they call 'Global Benchmarking Practices' – thus far they report on 247 different such indicators (2015). See Gisselquist for some guiding questions for evaluating governance indicators (2014).

## IV. Where does the data come from?

At the most general level you can distinguish between data that are from administrative sources and survey data. As already indicated, the survey data is a big mixed bag. Administrative data are more straightforward. There is a distinction between "survey data" and "administrative data." A survey is a specific tool the statistical office uses to collect responses from individual agents. Whether or not a statistical office is able to conduct surveys depends on its access to specific funding, as the normal budget allowance typically covers only the basic operation costs of the office. The administrative data are collected by public bodies to facilitate day-to-day governance and reflect the ambitions and extent of the activities of the state. The availability of data, which varies from country to country and according to the circumstances at a given time, determines the quality of the final estimates.

If you are studying administrative statistics, you are looking at the fingerprint of states, and their activities. There are a number of governance indicators out there, but looking at official statistics availability will give you a clear idea of general governance, as well as the types of governance the states are engaged in. I have made this point in studying the statistical offices and the official record in African states in the colonial and postcolonial period (Jerven, 2013a).

You can visit a well-stocked statistics reference library and bring a ruler. For a country like Botswana you will find meters and meters of publications, whereas the shelf on Equatorial Guinea will be slim indeed. Now consider data availability over time. In a country like Zambia, the 1980s and 1990s – or what William Easterly referred to as the "lost decades" in terms of economic performance – were indeed "lost" in the sense that statistical offices documented economic activities in a limited manner. You will find annual surveys and national account reports up and until 1973 after which there is a dearth of publications. As a final step you can look at what kind of data are produced at what times, and you will find that political (and donor) priority varies predictably over time. You will find labour surveys, transport surveys and industrial censuses for the 1960s and 1970s, whereas the only surveys of the informal sector are from the 1990s. Living Standard Measurement Surveys become the most important survey document as the World Bank shifts from growth to poverty, and since the 2000s the shelves will suddenly have publications on access to clean water, social indicators and reports on gender statistics.

The most recent Nobel Prize winner in Economics said that: "Politics is a danger to good data; but without politics data are unlikely to be good" (Deaton, 2015). By that, he meant that states collect statistics because they matter to states, so therefore the importance of statistics determines the quality and availability of statistics, but at the same time states may be tempted to tamper with the numbers, to put themselves in a favourable light. Today, the importance of numbers are undisputed, but it is equally clear that statistical capacity at the national levels is unequally distributed. It is normal to talk of the validity of statistics, and it is often interpreted as meaning whether the statistics are accurate or not. However, the root of the word valid, actually means power, and thus when we are talking about validity of statistics, we are talking of the social power, or the credibility and legitimacy of states.

While we demand these data to be available, and generally assume that they are, they do not exist for many. According to one report, six of the 49 countries in sub-Saharan Africa have never had a household survey and only 28 countries have been surveyed in the past 7 years (Chandy, 2013). A similar gap in coverage persists in surveys for social indicators, such as Multiple Indicator Cluster Surveys and Demographic and Health Surveys, and only about 60 countries in the world have vital registrations systems required to monitor basic trends in social indicators (Jerven, 2014a; Mikkelsen et al, 2015). In other words, monitoring of all indicators in all countries did not take place – partly due to insufficient funds and partly because these recording and surveying instruments are ill suited to developing countries.

Whether the data is primarily or typically collected from administrative or survey systems does vary from country to country, and as a general rule, in countries with weaker capacity in state administration data are necessarily drawn from survey sources rather than administrative sources. The objectivity of the data is generally believed to be higher in survey data. It has been well documented that in poor countries data on improvements in agricultural production, health and education tends be overstated in the administrative data (Jerven 2013).

Of the 60 MDG indicators in effect the majority of them require survey data. Some of them (like schooling, or health indicators such as mortality or number of births) are sometimes provided as administrative data in high income countries. But they require some kind of survey or census reference because they are phrased as 'proportion of…' which makes reference to a universal, valid population measure. In practice, administrative education, health data and civil statistics are drawn from medicals institutions, line ministries and official registered births and deaths. When schooling and health has limited reach, only a marginal share of the population are included in civil registries and only a small proportions of deaths and births are covered. The majority of the listed indicators in the MDGs and SDGs are resource intensive survey data, which countries in the bracket below $1500 GDP per capita will have a great difficulty in supplying without direct donor interest and funding.

In sum, a data user is well advised to take note of the fact that the quality of the observation varies substantially from source to source, and from country to country. Thus, training in handling data should be widened to include source criticism, and it is instructive to note that the international databases are very often not the primary source. Thus an understanding of the quality of data necessitates a visit to the statistical office, and a simple download from the World Bank database will not suffice.

## V. How good are the numbers?

Before questioning the quality of the numbers, we might want to introduce a typology of data quality. We are interested in three things. First, is there any underlying data? On the most fundamental level, we are interested in data availability. For better or worse, the dominant scholarly position is that some data is better than no data.[9] The word data means 'what is given' but often what we encounter in the databases does not satisfy this criteria. There are more gaps than real observations in the MDG database. When the World Bank announced new GDP figures for 2016, the median low income country has not yet completed their GDP estimate for 2014.[10] In 2015, the Secretary-General of the United Nations, Ban Ki-moon announced that we reached MDG1 (halve, between 1990 and 2015, the proportion of people whose income is less than $1.25 a day). The Secretary-General failed to mention that we will not have those data until sometime in 2017 or 2018, and that even then, we will be missing data for about a third of the countries. Moreover, we actually never knew anything about ratios of poverty in about half of the countries in the 1990s (Serajuddin et al, 2015). In sum: the first question you would be asking if you are interested in data quality would be: is there any?

It is not so obvious that the missing data is random. On the contrary. It is a well know truism that in surveys you are always missing out on the poorest and richest. This because neither Bill Gates nor the homeless, for very different reasons, tend to pick up the phone or answer the doorbell. At the global level, with countries as respondents, this may be equally or more serious, and you might also just be missing the lower end of the distribution. It seems safe to assume that we are likely to be missing more data points from countries that are so-called failed states, places that are embroiled in civil war or conflict, or places that simply do not have the resources to generate the statistics to take part. GDP growth may still be reported in some international databases, but the provenance of these may vary, and are at best described as estimates. Thus data availability matters.

---

[9] I reviewed the 1990s empirical growth literature for notes on data quality and found very few instances were data quality was questioned, but data availability was regularly lamented and reported (Jerven, 2014a).

[10] Based on the information collected in Jerven (2013). Pastor (2009) reports the same, reproduced in Appendix 1.

The second priority is of course that the data tells you anything. It is useful to think of this in terms of validity and reliability.  Reliability differs from validity in that a measure can have a predictable mismeasurement, so that it is incorrect, but predictably so. This mismeasurement would render the measure invalid, but the measure would still be reliable. We know for instance, that no GDP measure can be valid. That is, one could always argue that a definition of 'value added' can be changed, or adjust, improve or contest a data source. Measuring GDP is like standing on most bathroom scales: they may be wrong by a little or a lot. This matters less if what interests you is whether you are gaining or losing weight. However, if you are interested in comparing yourself with your friends or neighbours, accuracy matters a lot.

Thus, if the inaccuracy differs across time or space, the data are in effect incomparable. In terms of GDP per capita, if the level estimates are inaccurate but this inaccuracy was the same across time and between countries, the evidence could still be useful for comparison.  Unfortunately, as will be shown later, this is not the case. Therefore one has both validity and reliability issues with most development data.

In the following we will evaluate the quality of the key data for low income and lower middle income countries (as classified by the World Bank for 2016). This encompasses 82 countries and about 2,9 billion people.[11] 43 of the countries are found on the African continent, 7 in Latin America, 5 countries are small islands in the Pacific, while the remaining 27 countries are found on the Eurasian continent from Ukraine to Indonesia. It is implied that only this group has questionable statistics. Some low-income countries punch above their weight in terms of the quality of the data that they provide, while other countries provide less reliable statistics than we would expect from their income level alone.

[11] Afghanistan, Armenia, Bangladesh, Benin, Bhutan, Bolivia, Burkina Faso, Burundi, Cabo Verde, Cambodia, Cameroon, Central African Republic, Chad, Comoros, Congo, Dem. Rep., Congo, Rep., Côte d'Ivoire, Djibouti, Egypt, Arab Rep., El Salvador, Eritrea, Ethiopia, Gambia, Georgia, Ghana, Guatemala, Guinea, Guinea-Bissau, Guyana, Haiti, Honduras, India, Indonesia, Kenya, Kiribati, Korea, Dem Rep., Kosovo, Kyrgyz Republic, Lao PDR, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Micronesia, Fed. Sts., Moldova, Morocco, Mozambique, Myanmar, Nepal, Nicaragua, Niger, Nigeria, Pakistan, Papua New Guinea, Philippines, Rwanda, Samoa, São Tomé and Principe, Senegal, Sierra Leone, Solomon Islands, Somalia, South Sudan, Sri Lanka, Sudan, Swaziland, Syrian Arab Republic, Tajikistan, Tanzania, Timor-Leste, Togo, Uganda, Ukraine, Uzbekistan, Vanuatu, Vietnam, West Bank and Gaza, Yemen, Rep., Zambia and Zimbabwe.

There are a few indicators that are supposed to capture or indicate the quality of the official statistics. The most thorough framework is presented by the IMF and their Data Quality Assessment Framework (DQAF) that the IMF developed in the early 2000s in response to growing concerns about the quality of data provided to the IMF board. The DQAF serves as an assessment methodology, providing a structure for assessing data quality by comparing a country's statistical practices with best practices. In addition to a set of prerequisites for quality (such as the legal and institutional environment for data), the framework addresses five dimensions of data quality.[12] Metadata that are reported on the Dissemination Standards Bulletin Board also follow the structure and terminology described in the DQAF, and the DQAF provides the framework for the Data Module of the Reports Observance of Standards and Codes.[13] These reports are done by an IMF staff mission to the country, which collects information that is then organized according to the DQAF framework. The five-part structure contains two tiers of subcategories and ultimately comprises 50 different dimensions. The detailed assessment involves scoring the country on these 50 quality dimensions using the metrics of Observed, Largely Observed, Largely Not Observed, or Not Observed. A full data ROSC will conduct this scoring for the country's statistics in the real, external, fiscal, and monetary sectors.

The final report provide a detailed summary of many dimensions that together summarizes the inputs required to generate high quality statistics, but do not actually offer an evaluation of the quality of the numbers as such. The data ROSCs are resource intensive, and since 1999 there have been 111 data ROSCs. Only 43 of these have been in Low Income Countries (the last one conducted in 2009). Due to budgetary constraints, the ROSCs have almost been discontinued. Through these reports you can get a detailed summary of practices at the statistical offices a decade ago or so, but since very few countries have been examined twice it is futile to attempt to use this information to say something definite about a) the error margins in the data or b) how they may change over time.

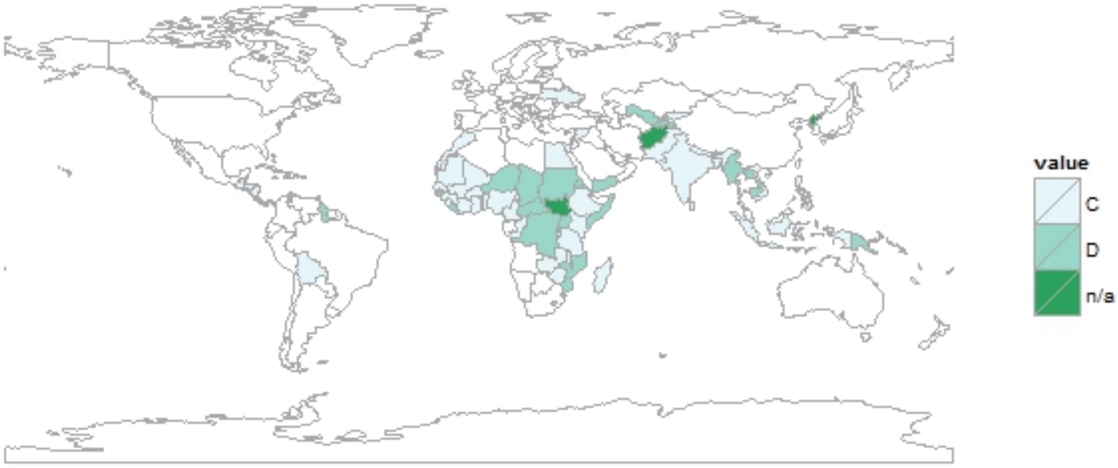
[12] 1. Assurances of integrity, 2. Methodological soundness 3. Accuracy and reliability 4. Serviceability 5. Accessibility.

[13] Variations of this framework are used as a framework for PARIS21 countries when they design national strategies for the development of statistics.

The IMF routinely collects economic information on member countries, through their Article IV surveillance reports. In these reports there is also a mandatory form (Statistical Issues Annex) where staff is to provide an assessment of the adequacy of data for surveillance. The report should highlight particular issues to note, while also containing a letter grade, as to whether the data is (A) adequate, (B) broadly adequate or (C) inadequate for surveillance. There are two important problems about this. The first is that there is no time, nor does the travelling economist necessarily have the time to assess data quality accurately. The second, is more serious. The economist collecting the data is the same who is assessing the data quality. It is also the same individual who writes the report to the board, using those very numbers to give a recommendation to the board. It would be a self-defeating act to declare that debt levels as a share of GDP are fine, but at the same time report that data on both are too weak to make that call. This concern has been raised in IMF reviews (2008) where "some upward bias in characterizations of adequacy" was found, while a more recent (IMF, 2012a) review concluded "… there may be some hesitancy by teams to use the 'C' classification."

Alternative guides to data quality was provided by The Penn World Tables. Up until version 6.1 it issued letter grades relating to the relative reliability of the estimates. Each country was assigned a grade of A, B, C, or D. Notably, none of the low or lower middle income countries scored above a C in this regard (Figure 1). For the 82 countries examined here, 44 of these were awarded a C grade, 27 D, and 13 were not included (see Appendix 2).

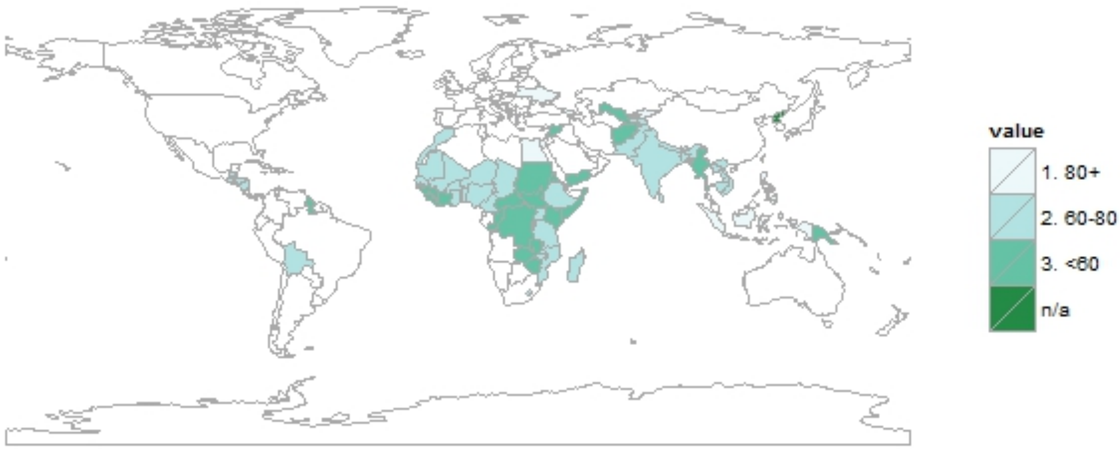
**Figure 1: Relative Reliability of Estimates in PWT 6.1**




The World Bank provides a more detailed approach, and ranks countries out of 100 in its Statistical Capacity Indicator.  Countries are scored on methodology, periodicity, and source data – as well as in overall terms. Figure2 reflects individual scores for our population of low-income countries, averaged out across 2010-2015. For the 81[14] countries being examined, the mean country score for the given period was 64.0. Low-income and lower-middle income countries scored an average of 57.7 and 65.6 respectively. The metadata are themselves harvested from the metadata that the countries are reporting to the IMF, but which are not always fully up to date (see also Appendix 3).

[14] No figures are given for North Korea.

**Figure 2: Statistical Capacity Average Scores (2010-2015)**




Data quality then, is hard to judge from afar, and it is difficult to give a precise assessment of the general level of data quality. To dig further into the issues of error margins we are here investigating the most important metrics of development: GDP, Poverty Headcounts, Purchasing Power Parities, Population and different indicators relating to MDGs and SDGs.

a. GDP

On November 5, 2010, Ghana Statistical Services announced new and revised GDP estimates. As a result, the estimate of the size of the economy was adjusted upward by over 60 percent suggesting that, in previous GDP estimates, economic activities worth about US$13 billion had been missed (Jerven, 2013b). While this change in GDP was exceptionally large, it did not turn out to be an isolated case. On April 7, 2014, the Nigerian Bureau of Statistics declared new GDP estimates. GDP was revised upward to $510 billion, an 89 percent increase from the old estimate (Jerven, 2014b).

The data mattered to investors, and therefore the financial press. In an editorial on October 28, 2013, the Financial Times wrote: “Reliable data are sorely needed. The

International Monetary Fund has warned that ‘the quality of basic economic statistics in sub-Saharan Africa . . . is often so poor that it can lead to serious misdiagnoses. In the past, similar problems have afflicted regions such as Latin America, the former Soviet Union and South East Asia.” (‘Africa at Dawn’, 2013)

The concurrent large and seemingly abrupt changes in GDP have led to a reconsideration of the quality of the data needed to evaluate basic trends in growth and poverty in LICs (Jerven, 2013a). At the same time, according to the African Development Bank, such substantial revisions have “understandably alarm[ed] many observers” (AfDB, 2013, p. 9), with the World Bank’s chief economist for Africa writing of “Africa’s Statistical Tragedy” (Deverajan, 2013; Jerven & Johnston, 2015d).

The GDP numbers that are reported ‘now’ in the databases are not official estimates, they are preliminary forecasts. Instrumental in creating the ‘Africa Rising’ narrative was a report by the Economist that said that seven out of the ten fastest growing countries in the world were in Africa (Jerven, 2014g). When *The Economist* was creating the ‘seven out of ten’ meme, it was using IMF forecasts made for 2011 to 2015. There is a difference between forecasted and actually measured growth. Studies show that forecasts for poor countries are likely to be overstating growth (International Monetary Fund Evaluation Office, 2015). Moreover, in poor countries where independent data production capacity may be low, forecasts may often stand in for real measured growth for a considerable time.

In a sample of African economies, the average reporting gap was found to be about a year and a half, and many countries are years behind (Jerven, 2013a). So when the IMF and the World Bank issue new numbers for the continent, they are based on guesstimates from a sample of countries. The average rates of growth of the reporting countries are extrapolated to the non-reporting countries. Because reporting countries (such as Rwanda and Ethiopia) may be doing better than non-reporting countries (such as Somalia and Guinea-Bissau), growth is systematically overstated. Finally, many are questioning the validity of such high reported growth rates. In the case of Ethiopia there has been an open disagreement about the growth between the authorities and the IMF. Ethiopia claims to be in double digits whereas the IMF reasons the growth is lower. Corroborating the IMF story, Geda and Yimer (2014) researched the validity of Ethiopia’s national accounts, and the official report of eleven percent per annum growth from 2002/03 to 2010/11. Motivated by a “marked absence of structural transformation in the economy where agriculture still contributes nearly 50 percent of the GDP” with “no evidence of intensification of agriculture”, the authors conclude that the correct annual growth rate is in the range of 5-8 percent – still commendable in their opinion. A variety of methods are employed in this endeavor including a comparison of national accounts figures to independent institutions (such as the National Bank of Ethiopia); adopting Alemayehu et al’s (2004) “growth accounting” and calculation of the country’s maximum growth potential; extrapolating backwards via finance’s “doubling rule”; and examining the supposed sources of the reported growth by looking at total factor productivity.

The quality of the national income estimates is thus a result of the combined quality of the activities at the statistical office. The national accounts division depends on data that are produced in different parts of the statistical office – particularly for data on population, agricultural and industrial production and information on prices. The supply of data from these sub-divisions is subject to the manpower and funds available for data collection and processing. Frequently, the statistical offices rely on data made available from other public and private bodies. For example, the agricultural data will typically come from the equivalent of the ministry of agriculture. In some sectors that are dominated by a few large operators, such as construction, mining, electricity, water, finance communications, transport, the office will depend on the supply of data from these private or public entities. Here, the GDP aggregate will rely on a mix of ‘survey data’ and ‘administrative data’

The basic questions are whether the statistical office has any data, how good they are and what the national accountants are to do when data are missing. The first step in the aggregation process is to make a baseline estimate or a benchmark year. The best instrument here is a census. This can be a census of the population, the agricultural production or the transport sector. If a census is absent, a survey may be available. A survey contains some information about a sample of the total. If there ever was a census, then you can aggregate these results, assuming that the sample is representative. If there is no total to relate the survey to, the statistician will have to make a guesstimate, literally making up the missing information without any official guidelines. Often there is no data. When there is no level data the compilers have to rely on estimation by proxy, or assumed relationships. A classic example is when one has no data on food production and then assumes a per capita calorific intake which is multiplied by a guess of the farming population. Data are usually missing for parts of the service sector and a common method is to assume a proportional relationship with production of other physical goods.

When a level estimate for a given year has been reached, the wealth of the nation is measured. The next step is to measure economic growth, so that the progress of the nation can be monitored. It would be easy to get the impression that this would simply entail aggregating all available data once more and comparing the current year with the previous. The way this is done in practice is quite different. The level estimates for the individual sector are already made and form the basic starting point. In some sectors, such as government expenditures and turnover for larger businesses, one is able to compare the total for one year with another, but for large parts of the economy one usually relies on so called 'performance indicators' or 'proxies'. The annual data collected from public bodies and private businesses are utilized. These are supplemented by data on exports and imports. A typical example is the use of cement production and/or imports as a proxy for growth in the construction sector,[15] the number of new official licenses for transport sectors and reliance on population growth for sectors where little adequate data are available.

There is a basic distinction between making a level estimate, referred to as a base year, and that of estimating change. One can think of it in terms of an object's weight. The weight may be inaccurate, and show the entity measured to be too heavy or too light. If the degree of inaccuracy was reliable, it would not matter much for measuring change. That is: even if a weight shows you to be too heavy, if the weight is equally skewed the next year, you would at least know with accuracy how much weight you have gained or lost. There is one mathematical caveat to this: since change is measured in percentage, you will appear to be gaining weight at a faster rate if the weight showed you to be lighter than you really were. Following from these principles one could expect the following: the more underestimated the level of GDP, the more overestimated growth would be. This is mathematically true when one is measuring the weight of a

[15] Despite the importance of construction sector, it remains poorly measured. For an example from Kenya see Kenya K'Akumu (2007).

person, or the number of oranges in a bag. It is not true when one is measuring GDP, because the GDP is a composite index with a base year.

The base year estimate is of crucial importance. It determines the proportional shares of different sectors of the economy. The issue resulting here is generally referred the as the 'index number problem' (Feinstein, 2002, Appendix B4). The size of each individual sector determines the impact that the growth in one sector has on the aggregate growth. In order to measure 'real' economic growth, the economy will be accounted for in the base year's prices. This is done by either deflating a sector with a measure of inflation often applied to the service sectors, or by expressing output in the base year prices directly, as in multiplying the physical output in mining or agriculture with the prices obtained in the base year. Generally speaking, the less 'normal' (as for instance a year of drought) and the older the base year, the more misleading the growth series will be. This also means that if one part of the economy is underestimated, its contribution to aggregate growth will also be underestimated.

One way of gauging the quality of the GDP estimates, or at least getting an idea of whether the levels are roughly in line with those from other countries may be the age of the benchmark year.

**Figure 3: National Accounts Base Year**

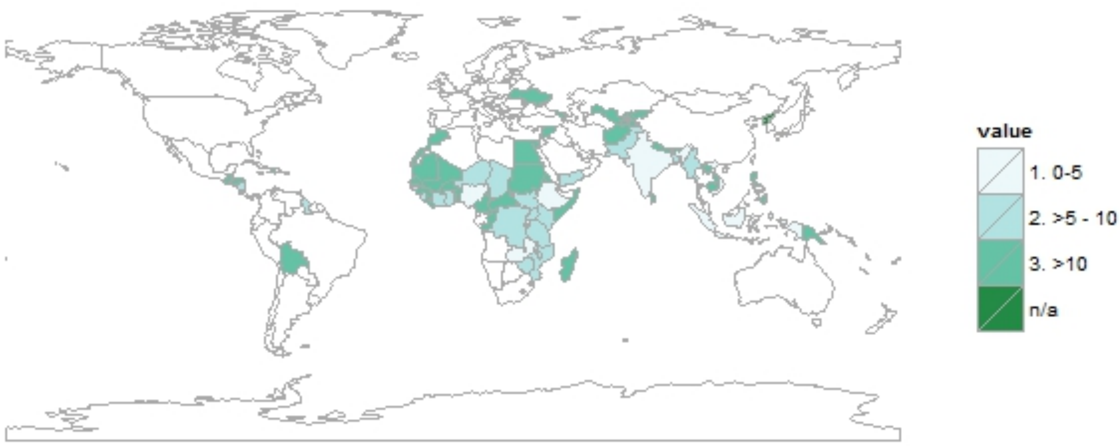
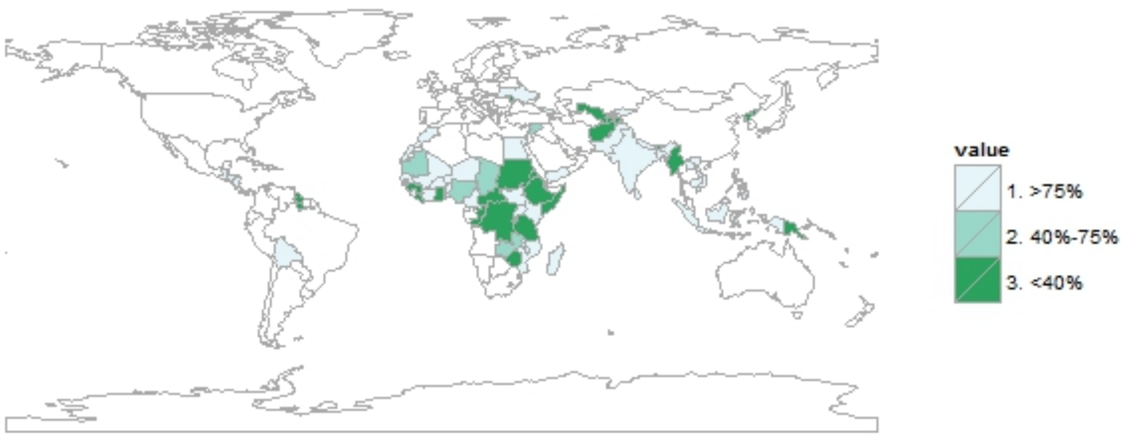


Figure 3 maps countries based on the base year for their national accounts, as per the World Bank. Notably, the average base year is over 13 years old (as of 2015), with a standard deviation of 6.8.[16]

Moreover, in many cases, on the basis of the information in the databases we do not even know whether the estimate reflects what the country actually is reporting GDP to be or whether it is an estimate from other sources. Thus, while out-dated base years are a problem, so too is the fact that many countries have been sluggish – or have completely given up – in providing GDP estimates to the IMF in a timely fashion (see Jerven, 2013 and Ward, 2004 for the history of GDP reporting). The World Bank may be providing data for 2015, but most developing countries are still calculating GDP for 2012 or 2013. In the meantime the databases are populated by forecasts and through other methods of gap filling.  There are exceptions. Rather than guesstimating based on proximate trends, the IMF's International Financial Statistics database only publishes GDP estimates that actually are reported by national statistics offices to them.

**Figure 4: GDP statistics availability in IFS**




As of January 2016, the most recent year for which GDP statistics are reported is 2014. For the period from 2001-2014, of the 82 countries in our sample, GDP statistics were

[16] North Korea is excluded, due to missing information

reported at an average of 57% of the time. Most alarming is that 23 of the 82 countries have no GDP statistics reflected in the IFS database at all (see Appendix 4). Figure 4 reports the availability of GDP data in the IFS as percentage of possible country year observations from 2001 to 2014.

Of course, it is important to underline that whether the benchmark year is old or not, and whether the IMF or the World Bank accepts the country forecasts and estimates, the GDP metric is still misleading. It is fundamentally flawed in measuring 'development' as such, for many reasons.[17] But even if you are just interested in economic growth per se, only parts of an economy are actually recorded from year to year. Though different allowances can be made to make up for the missing data in the informal sector in level estimates, year-to-year growth is driven by the visible parts of the economy. Thus growth, on paper, is driven by easily observed items such as exports and foreign direct investment. Meanwhile, important sectors that may be moving less quickly – such as food production – are under-observed, and thus we may overstate growth in periods of expansion and overstate economic decline in times when external sectors are not doing so well.

This creates serious problems for those who are interested in economic growth, and furthermore the changes in benchmarks are unequally reflected in the different international databases, and thus reported growth rates can also vary. Derek Blades reported that while there was high error attached to GDP level estimates, estimates of growth were probably better, but he warned that estimates of year-to-year variation should be treated with extreme caution. He concluded, "Thus an estimated year-to-year increase of 3 percent might mean anything from no growth at all to an increase of 6 percent" (Blades, 1980: 60). Further research on growth rates reported on a subset of African economies confirms this concern, and notes that at times the error margin is even larger.[18] This makes it difficult to associate policy change with changes in growth (Jerven 2014a) and furthermore since growth is poorly measured one would hesitate to equate it with a rise in living standards without corroborating evidence. The considered prime source of such evidence would be measures of poverty rates and changes in levels of poverty.

---

[17] A point not worth laboring here, as it is already made even-handedly by Stiglitz, Sen & Fitoussi (2010); and Coyle (2014) with a more critical approach by Fioramonti (2013).
[18] See Jerven 2010a, 2010b, 2011a, 2011b.

b. Poverty

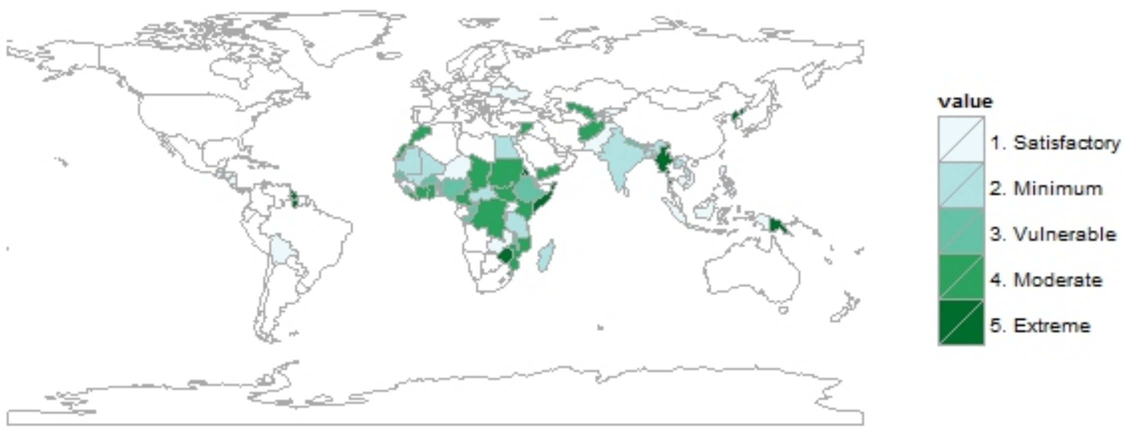
The focus will be on counting poor people as suggested by the World Bank and the Living Standard Measurement Surveys implemented sporadically in various places in poor countries since the 1980s, and then with renewed strength since the adaptation of the 2000 Millennium Development Goal of halving the proportion of people in poor countries living in extreme poverty since 1990. It is arguably the most visible global indicator of success and failure in international development. According to a study by World Bank researchers, for the period 1990 to 1999, the World Bank only had access to adequate data in 43 of the 150 countries it was purportedly monitoring poverty (Serajuddin et al 2015).

In Figure 5 these categories of what Serajuddin et al calls 'data deprivation' is replicated, with the figures updated to the year 2012. The Figure reflects the poverty data deprivation in the developing world by looking at the frequency of poverty surveys (as listed on povcal.net) over a ten year period. The starting point is obvious. We need two data points in order to analyze a trend. Accordingly, countries are coded in one of five categories:

1. *Extreme Data Deprivation:* If no poverty data point in the ten year period
2. *Moderate Data Deprivation:* If only one data point in the ten year period
3. *Vulnerable* to *Data Deprivation:* If two data points, but separated by more than 5 years
4. *Minimum Requirements for Data Needs:* Two data points in ten years, 5 years apart
5. *Satisfactory for Data Needs:* Three or more data points in the ten year period

Overall, according to these categories, 14 out of 82 developing countries are graded as extreme and 24 as moderate. No data or just one point does not equate to having enough information to say anything about a trend. So it would perhaps be more fair to say that for 38 of the countries, the data is not there, or simply inadequate. There are 10 countries that have two data points within a decade, but they are more than 5 years apart, not enough perhaps, to be confident about a trend in poverty. Thus, depending on one's judgement, the poverty data are inadequate for slightly less or slightly more than half of the low income countries in the world (see Appendix 5).

**Figure 5 – Poverty Data Deprivation in the Developing World (2003-2012)**




It is not only about data availability. These surveys are a method to get to know about expenditures and monetary poverty, but they depend on a range of factors for bringing accurate results. One, seemingly inescapable, weakness it that the surveys depend on the definition of the household being applicable (Guyer 2004, Randall and Coast, 2014). Other obvious difficulties relates to seasonality (when is the survey conducted), fatigue (how long is the survey), and ability and willingness to recall correctly expenditures on various goods and services in a certain time period. It is hard to gauge how this influences results, but survey design has been found to determine results. Gibson et al (2015) build on Beegle et al's (2012) work by looking at household consumption in Tanzania. They find that variables such as survey length, data capture (diary vs. recall), frequency of researcher visit, whether measurement is an individual or household task, and recall period all had a statistically significant impact on the measurement error of the survey. Kilic and Sohnesen (2015) add further cause for concern by finding that respondents answer identical questions in different ways, depending on how long a questionnaire is. Specifically, their work in Malawi finds that just the length of the survey itself can impact poverty rates anywhere from 3 to 7 percentage points. A study in El Salvador from 1994 showed that more detailed questions on consumption resulted in an estimate of mean household consumption that was 31% higher than that from the short version of the questionnaire (Jolliffe, 2001). The same findings have been confirmed in labour and agricultural surveys that show very different level results depending on survey method and time of survey (Jerven and Johnston 2015). As one would recall, this is a validity problem. It is hard to get the level right, but the validity problem becomes a reliability problem if you make comparisons over time or if different surveys are implemented over time in the same country.

c. Purchasing Power Parity

None of the GDP or Poverty data would make full sense if there was no attempt to express GDP per capita or poverty lines in international dollars, that is, controlling for the fact that one dollar has a different purchasing power in different locations. The national income data are aggregated in local currencies. The first step to take towards making the figures comparable is to use foreign exchange rates in order to express one country's income in the currency of another. This does not take care of the problem of differences in domestic prices on non-tradable goods, however, or other factors that cause a divergence in purchasing power parity (Taylor and Taylor, 2004). To achieve purchasing power parity (PPP) one needs to adjust for the fact that one dollar goes a lot further in Ethiopia than it does in Canada. This entails a complicated process of collecting prices and then determining a basket of goods and services to weight the individual prices (Deaton and Heston, 2010). Price data and baskets are irregularly updated under the auspices of the International Comparison Program.[19] Lack of comparable information may introduce bias in these comparisons; it may well be the case that in Ethiopia, for example, prices are relatively more readily available for urban areas than for rural areas, and furthermore that there may be better price data on imported goods than on domestically produced subsistence goods. If the urban data on imported goods are weighted too heavily in the consumption baskets, the PPP adjusted GDP figures will understate the living standard in Ethiopia.

Agreeing on a poverty line and adjusting national surveys to express living standards in purchasing power parities is a daunting task. Angus Deaton (2015) recently called it 'a statistical problem from hell'. And it is not only a methodological and political problem as, practically, it is both expensive and time consuming to survey households and aggregate these to comparable numbers of the poor.

[19] For a view on this problem in longer historical perspective, see Jerven 2012.

The International Comparison Program (ICP) is a global statistical partnership designed to calculate purchasing power parities across the world. Beginning in 1970, there have been eight ICP rounds, with the most recent being in 2011. Table II reflects the years and participation in these rounds while Figure 6 reflects the participation of the developing world.

**Figure 6: Rounds of the ICP Participation**

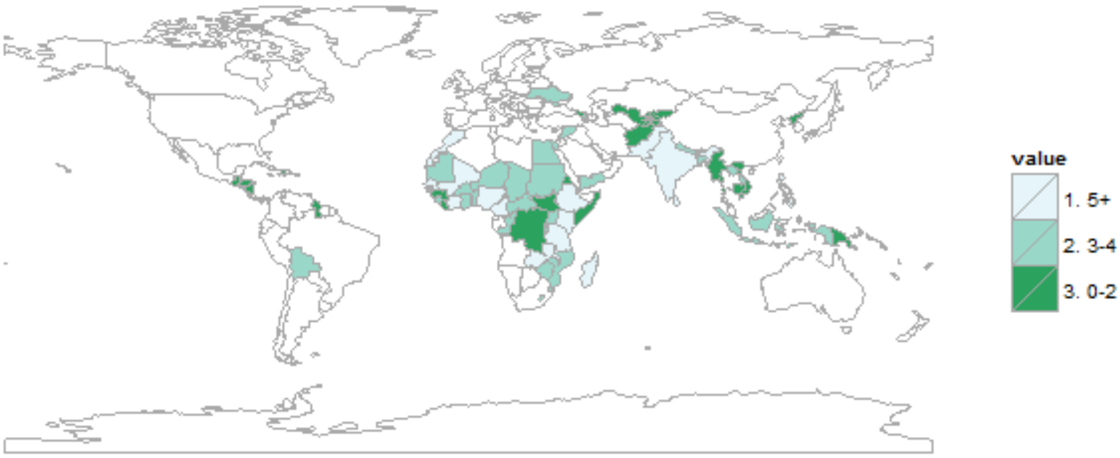



**Table 2: Rounds of the IMF's International Comparison Program (ICP)[20]**

| | 1970 | 1973 | 1975 | 1980 | 1985 | 1993 | 2005 | 2011 |
|---|---|---|---|---|---|---|---|---|
| Participants | 10 | 16 | 34 | 60 | 64 | 117 | 146 | 199 |
| Developing Countries[21] | 2 | 3 | 8 | 22 | 25 | 46 | 60 | 71 |

Deaton (2010) highlights that these PPP rebasings are more than just exercises, nor just a basic update of our knowledge. New data seems to contradict the old data. Using PPP values from the then most recent three ICP rounds (1985, 1993, and 2005), he looked at estimates for the global poverty headcount in 1993. This gave three different estimates for poverty headcount – one for each PPP date. A selection of these results are

[20] 1993 participants are taken from UN Statistics <http://unstats.un.org/unsd/mdg/SeriesDetail.aspx?srid=699> (21 January 2016).
[21] See Appendix 6.

reproduced in Table 3. The rebasing of PPPs had the statistical effect of increasing poverty measurements (for the same actual year) from 29.4% (1985 PPP) to 39.2% (2005 PPP); representing half-a-billion newly-qualified poor. Complicating the situation is that these effects were not uniform across regions. For example, Sub-Saharan Africa was found to have 17 million more living in poverty and East Asia/Pacific doubled from roughly 25 million to 50 million. On the opposite end of the spectrum, Latin America more than halved its poverty headcount from 23.5 (1985 PPP) to 10.1 (2005).

**Table 3 – Global Poverty Headcount in 1993, According to Different PPP Dates**

| PPP | | Headcount (millions) | | | |
|---|---|---|---|---|---|
| Year | Global Poverty % | Global | Sub-Saharan Africa | E. Asia / Pacific | L. America |
| 1985 | 29.4% | 1,350 | 39.1 | 26.0 | 23.5 |
| 1993 | 28.2% | 1,304 | 49.7 | 25.2 | 15.3 |
| 2005 | 39.2% | 1,799 | 56.9 | 50.8 | 10.1 |

Source: Deaton (2010)

There is little way of guaranteeing the actual provenance of these price data. [22] As has already been pointed out, these offices are sometimes suffering a reporting burden to the extent that they have stopped reporting real sector statistics to the IMF International Financial Statistics, and may have Consumer Price Indexes that are severely out of date, or are only collected in urban areas, so if the questionnaire reported to the ICP round is a bit lacking it would hardly be a surprise.

Surprise, was the word that summarized much of the development community when the new ICP results came out in May 2015. Laurence Chandy and Homi Kharas called it a statistical earthquake (2015), whereas researchers at the Centre of Global Development used the new PPP data to calculate that 'Global Absolute Poverty Fell by Almost Half on Tuesday' (Dykstra et al 2015). In response, the World Bank officials responsible resorted to authority, and while recognizing the entertainment value of

[22] I attended a conference where the experience of collecting such price data was discussed, and while the latest ICP round give the impression of full participation that has come at a cost of comparability and accuracy. An ICP participant recalled receiving an excel sheet with prices from an undisclosed country, and was puzzled by the high variation in the prices, and that the variation was random. An informal query was sent back to the country office for clarification, and quite shortly thereafter the ICP staff received a much tidier distribution of prices. This was shared on 2011 IARIW-SSA Conference on Measuring National Income, Wealth, Poverty, and Inequality in African Countries, Cape Town, South Africa, Sept. 28 - Oct. 1, 2011.

millions jumping across the poverty line overnight, still warned about the pitfalls of "Playing with and Understanding Purchasing Power Parities"(Basu, 2015).[23]

Thus, when it was announced that we met the MDG1 of halving world poverty, it was a classic case of the PR department being one step ahead of the knowledge department. The goal was to halve the share of the population living in extreme poverty in 1990. Yet, the data presented here tells us rather clearly that we know very little about poverty in 1990. We are talking about several hundred millions of people in the error margin. Moreover, the data since 1990 is sporadic. In the past decade we have survey data on about half of the countries to say something about trends. Finally, the 2015 data are not ready yet. Data since 2012 is relying on extrapolation from the not so reliable or missing GDP data.

d. Population Statistics

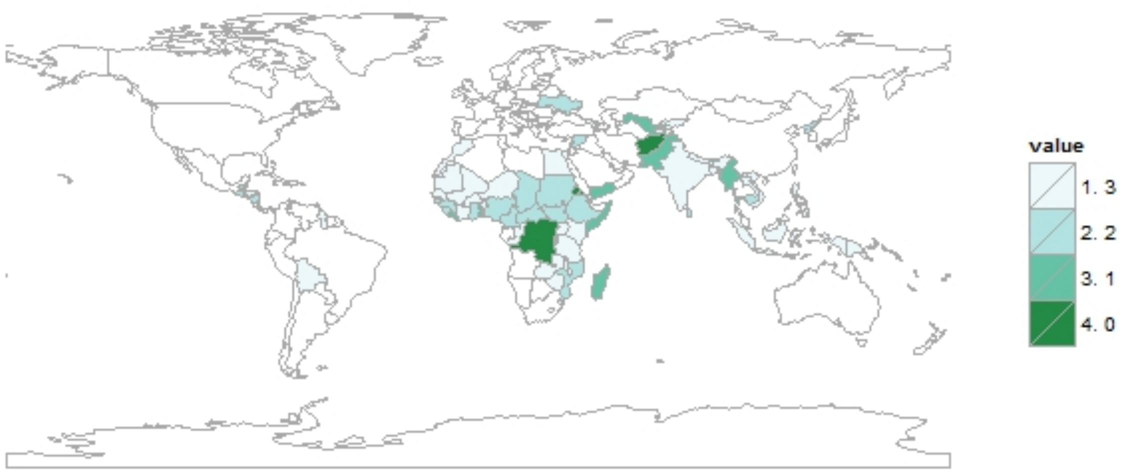
The accuracy of the count of a population depends on the state's capacity to undertake the population census as well as the incentives for the enumerators and the respondents. Take the example of the history of census-taking in Nigeria. The 1952/53 Nigerian census produced a total population of 31.5 million. A decade later, the Nigerian Ministry of Health used this figure to calculate a total population of 36.5 million using a 2 percent rate of growth between 1952 and 1962 (Jerven, 2013). Thus, it was with great surprise that the 1962 census gave a final count of 45.3 million and even more so when a count one year later brought the total to 55.7 million (Caldwell and Okonjo, 1968). Which count(s) should we trust? The 1952/53 census is believed to have suffered from a response bias that stemmed from a general fear that the data would be used for taxation purposes which made native authorities reluctant to cooperate. However, in 1963 the political situation was reversed. In newly independent Nigeria it was expected that transfers from the central government (such as funds allocated for schooling, health, and infrastructure) would be positively dependent on population numbers. Through

[23] The comment field under the Centre of Global Development post was particularly entertaining, with one commenter reflecting on a the response from Martin Ravillion, long in charge of calculating the poverty headcount thus, "Measuring absolute poverty is difficult and Martin Ravallion is apparently the only person on the planet capable of doing it properly. Kudos to him and shame on the rest of you for even trying."

research evaluating secondary material, the demographic structure and the political economy of such counting problems can be resolved to some extent.

In a recent report, *The Economist* ('Population, We happy few', 2015) claimed that Nigeria's population is massively exaggerated, by as much as 20 million. Nigeria is undoubtedly the odd one out here, but it should be noted that within the study of the historical demography of Africa the population estimates in the colonial period are contested in the order of plus-minus 50 percent. The uncertainty between census counts and projections of populations are still felt today. Angola's 2014 census found that the actual population of the country was 24.3 million, more than 25 percent higher than existing estimates (''Angola has a population of 24.3 million people', 2014). This was followed closely by censuses in Uganda and Guinea that revised population figures downward from 36.5 million to 34.9 million (Olukya, 2014; Musoke, 2014; Kalungi, 2013, World Development Indicators, 2015) and from 12.28 million to 10.5 million, respectively ('Décret : les résultats du recensement général de la population guinéenne publiés', 2015; World Development Indicators, 2015). Errors can be committed in both directions.

Figure 7 – Number of Last Three Census Rounds Participated In (1985-2014)



The challenge here is that census data is not collected as frequently as one would hope for in developing countries. Figure7 reflects how many of the previous 3 census rounds (1985-1994, 1995-2004, 2005-2014) developing countries have participated in, per UN Statistics. While the average value of 2.28 is encouraging, that 5 and 11 countries participated in 0 rounds and 1 round respectively highlights that there are limiations to our knowledge about the denominator for the Low Income Countries (Appendix 7).

### e. MDGs and SDGs

Currently the international development community has embraced the idea of 'evidence based policy'. Related to it are the principles of 'results based management' that have inspired the development community to set out quantifiable targets such as the aforementioned Millennium Development Goals (Black and White, 2013). This has again put the issue of statistical capacity of poorer countries on the policy agenda. The eight goals are supported by 21 targets and 60 indicators, thus encompassing most aspects of economic development. Vandemoortele claims that statistics have been abused to fabricate evidence of success in the case of the MDGs, and that the use of the quantitative targets has furthered the one-dimensional view of development and this process has strengthened the "money-metric and donor-centric view of development" (Vandemoortele, 2011). Sanga argued specifically regarding the MDGs that: "a major weakness is the assumption that data would be available. Countries have been struggling to build their capacity to collect, process and disseminate the requisite data" (Sanga, 2011).

In some cases the monitoring demands of the MDGs, have also meant a windfall of economic resources for the statistical offices. National accounts divisions have complained that this means that staff from already undermanned divisions are pulled to sections where data are collected for the MDG indicators. National stakeholders, such as the central banks, have lamented that they suspect that the quality of the important economic growth data has been decreasing. It has been observed that, as a result of more resources for data collection, analysis and dissemination have suffered. These concerns have been echoed by representatives from the IMF and World Bank. The concern was that the limited capacity of the statistical offices was further constrained by the Millennium Development Goals agenda.

The response from the national accounts divisions, the statistical offices, international and national stakeholders is clear (Jerven, 2013). The pressure currently put on statistical offices is not matched by their capacity. A discussion paper by Gonzalo Duenas Alvarez, et al (2011), provides a listing of all the available data relating to 12 MDG targets from 1990 to 2009, for each sub-Saharan African country. The data availability picture is a mixed one: 9 countries have data at least as recent as 2005 for all but one of the targets (Liberia is the only country with recent data for all targets) and most countries have at least some data over the time period for all but one target. Somalia and Sudan have no data at all, and it is notable that the poverty data consistently are where we find the least recent observations. Note that this only surveys a subset of 12 MDG indicators and the data availability situation for all 48 indicators provides a more pessimistic picture.

There is not room to evaluate the reliability and availability of all of the MDGs here. The MDG reports have tended to be written 'as if' we have annual data on these indicators, but the fact is that we do not. At the launch of the 2014 Millennium Development Goals, Keiko Osaki-Tomita, chief of the demographic and social statistics branch of the UN's Statistics Division noted that for many of the indicators, including poverty the latest data was for 2010 (Scidevnet, 2014). The statistical offices will struggle to cope with increased data demands that will arise after 2015, with the new post-2015 development agenda and the Sustainable Development Goals. If we would multiply 60 indicators with the number of low and lower middle-income countries (82) and the number of years (25) we would end up with a very high enumerator (123,000). There is simply no practical way that this data gap could be met with credible surveys on progress. If you multiply it with a typical annual survey cost of 1 million USD dollar, you are looking at a total survey cost of just about the total annual official development assistance budget. This does not even include basic census data (without which the survey data are meaningless) and the cost of maintaining basic statistical infrastructure for facilitating collection and dissemination (which would crumble under such a survey demand).

There are more gaps than real observations in the MDG indicator database and many of those observations that are actually contained in the database are of dubious quality (Jerven, 2013). If we did not even ask for annual data, but accepted data every 5 years on most indicators the best possible cost estimates indicate that if the previous MDG agenda would have been measured it would have cost about $28 billion. Yet, as we know there were gaps in the data and many indicators were never properly measured between 1990 and 2015. A future agenda with 169 targets has an estimated cost that is higher than the total annual spent on official development assistance globally (Jerven, 2014).

In order to assess data availability, the IMF DSBB's Enhanced General Data Dissemination System (e-GDDS) was analyzed (see Appendix 8). This database provides insight into the source data and statistical practices behind economic, financial, and socio-demographic data. Under the assumption that health and education data is required to measure improvements in health and education measures (measures that feature prominently in the MDGs and SDGs), these country reports are analyzed for the quality of their health and education statistics .

While this table reflects a range of quality, the IMF has no information on the health statistics of ten countries and education statistics of eleven out of the total population of 31 countries. These figures increase to 23 and 25 respectively if we assume that information that is 10 years old is likely to be outdated. Between this extreme and the (supposed) gold standard of annual collection, we see a variety of other practices (including countries that rely on censuses – conducted every ten years – for all of their information). In case of agricultural and labour statistics, which are supposed to be part of the SDGs the data periodicity is similar or worse.

There is paucity of actual observation. In addition, there is a political economy of measuring that brings those observations into doubt. Welle (2014) highlights the impact that powerful actors have on the monitoring of water access, part of the MDGs, in a region of Ethiopia. Depending on the method of calculation, water access was found to range anywhere from 24% to 54% in the same region. Her work points out how the employment of different methods of data collection and calculation led to divergent portrays of access to water. Welle draws the conclusion that individual actors in positions of power can manipulate these processes to reach preferential conclusions (for example, increased access to water after a specific policy intervention), a condition labelled as "performing on monitoring".

In a similar tone, Sandefur and Glassman (2015) observe discrepancies between administrative data and independent household surveys. Specifically, they find that official statistics systematically exaggerate progress on health and education indicators and point to two causes for this problem. Governments are found to be misled by false information that is being reported from the frontlines, and to then inflate these numbers when reporting them to international donors. At both stages, the motivator is suggested to be funding, as the frontlines look to secure funding from the government while the government seeks to secure funding from international donors. Across their sample of African countries, a gap of 3.1 percentage points in change in enrollment rates is identified between administrative data and sample survey data. The discrepancy was as high as 21.4 percentage points in the case of Kenya, 1998-2003 (see Appendix 9).

## VI. How to do and not do development by numbers

So how good are the development numbers? To summarize the previous sections discussing the availability and reliability of the most important numbers in use in the study of economic development I have provided a table below.[24] The table below gives you a grade of the databases on GDP, Poverty, PPP and MDGs/SDGs summarizing the two key variables. How much data is missing (availability) and what is the likely upper and lower bounds (reliability)? The Grades are A (1-0.8), B (0.79-0.6), C (0.6 or less). There is no obvious justification for the grading categories. If 'A' was determined to be from 1 to 0.9 we would have no 'A' grades to give. So in a sense, I am grading to the curve, not to an absolute and perfect standard. The purpose is to grade the databases, and to create a summary statistic of the combination of availability and reliability. The rationale being that those are the two most important elements of data quality.

Imagine that we had perfect data availability for GDP for the European Union, and further imagine that studies show that GDP, under current definitions, is reliable to a variation of +/- 5 percent from country to country within the EU. Thus, on availability, the GDP metric would score 100 percent (1), and on reliability 95 percent (0.95). The summary statistic is the square root of the product of the two, which in this case would be 97 percent, and thus the grade 'A' would have been applied.

On the lower end of the scale we could look at Nigeria, where GDP growth real rates have not been reported to the IMF since 2003. The GDP levels were adjusted upwards recently by 89 percent. Hence, we would estimate something like 10 percent availability and +/- 80 percent reliability, and thus ending up with a letter grade safely in the C bracket. The average grades end up somewhere in the middle of these two extreme examples. The availability scoring is taken directly from the data presented in the appendices. For reliability, I have used judgement based on the many research findings on the accuracy of level estimates. In the case of GDP it is suggested that any GDP estimate from a low income country could be within a +/- 40 percent error band (as suggested by Blades 1980, and Jerven 2010c, and further suggested to be higher by Jerven 2013a). For poverty levels, plus minus 20 percent is suggested. For PPP the

[24] Country level measures of availability and reliability are reflected in Appendix 11, along with the methodology used to calculate them.

reliability scoring is taken from the Deaton (2010) exercise. The population error margin is similarly taken as a judgement based on reading of the literature evaluating population censuses, and results comparing projections and new census counts (Frankema and Jerven, 2014). The MDGs and SDGs margin of error is judged to be +/- 20 percent, although some the data reported here suggest that the margins can much higher in individual cases. The availability scoring for each country is reported in appendix table 11-

**Table 4: Grading the datasets.**

| Data type | Availability | Reliability | Quality | Grade |
|---|---|---|---|---|
| GDP | 0,56 | 0,60 | 0,58 | C |
| Poverty | 0,61 | 0,80 | 0,70 | B |
| PPP | 0,36 | 0,63 | 0,48 | C |
| Population | 0,76 | 0,90 | 0,82 | A |
| MDG | 0,50 | 0,80 | 0,63 | B |

It makes the point that on average, the data availability is weak. We are missing data just a little bit less than half of the time. One some level, validity is always a guessing game. These are not items that can actually correctly be measured, but the idea here is to provide a summary impression of the error range in each direction. So, if someone tells you that GDP per capita in one country is 1000 dollars and another it is 1200 dollars you should not, on average, be sure that that is a meaningful difference. In fact, only if the difference is in the order of 1000 in comparison to 600 or 1000 to 1400 you are approaching an area where you could safely say that there is a difference, though recent examples from Nigeria and Ghana illustrate that you can still be mistaken.

The scores are average scores for the countries in the lower third of the income classification provided by the World Bank, as has been painstakingly pointed out before, the average may be misleading. Thus, you will have countries where data availability is 'adequate', and while the poverty survey is generally reliable only to the +/- 20 percent grade, that would still leave a country that surveys poverty every fifth year getting the average score in the letter A grade. The aggregate picture is dismal, and reflects that if you pick a country at random you are not likely to get high quality information on poverty, growth or progress on MDGs.

There is ample evidence that the MDG agenda has already stretched statistical capacity and strained statistical offices in poor countries (Jerven 2013). Or, as it was summarized by Richard Manning, formerly of DAC-OECD in a DIIS report (Manning 2009, 38):

*It is not clear that the expanding number of surveys and data collection exercises has had a positive and sustainable impact on local capacity. It is quite possible that we are in fact seeing a growing mismatch between the multiple demands for monitoring and the ability of local systems to generate credible data. There is a danger that an 'MDG Results Industry' could consume a lot of resources to rather little effect.*

The proposed new post-2015 list is likely to stretch this gap even further. The post-2015 MDG debate has so far been dominated by what goals and targets are desired and as of yet less discussion about what can be realistically measured, what kind of indicators might be needed and even less concerning who should pay for the measurement. One could take the view that right now the concern should not be 'how much does it cost', but rather first determine 'what do we need', and then later on figuring out 'how we pay for it'. I strongly suspect that the latter will be the ad hoc approach generally taken, but I would not recommend such an approach. The cost of monitoring should be taken into account. It is not the case that all increases in measurement activities are improvements in overall statistical capacity. Provision of data has opportunity costs, and provision of data further has behavioral implications.

One might read this paper as a challenge to future empirical work.  It is, but it is neither an unfair challenge, nor is it an unsurmountable one. There is a lack of critical scholarship on numbers. I noted in section IV that there is a rise in use of numbers for decisions, for research, and also for public consumption. In social media the so-called 'data visualization' is meant project wisdom and insight, and is a clear indication of our need to have shorthand knowledge to make sense of the globalized world. But without a clear idea of where the data come from or how they are generated, such knowledge about the world may indeed be very shallow. There is now a rising literature, matching the rise of use of indicators and numbers in global governance and public life. Thus, I hope that this paper may serve as a basis for future interrogations of the knowledge and governance challenges in doing development by numbers.

## References.

Abdelal, R. & Blyth, M., 2015. 'Just who put you in charge? We did: CRAs and the politics of ratings', in Ranking the World: Grading States as a Tool of Global Governance, eds. A. Cooley & J. Snyder. Cambridge University Press: Cambridge.

'Africa at Dawn', Financial Times, 28 October, 2013 <http://www.ft.com/intl/cms/s/0/3bb87da0-3fce-11e3-8882-00144feabdc0.html#axzz3x45CvcVX> (12 January 2016).

African Development Bank, Statistics Department, 2013. Situational Analysis of Economic Statistics in Africa: Special Focus on GDP Measurement <www.afdb.org/fileadmin/uploads/afdb/Documents/Publications/Economic%20Brief%20-%20Situational%20Analysis%20of%20the%20Reliability%20of%20Economic%20Statistics%20in%20Africa-%20Special%20Focus%20on%20GDP%20Measurement.pdf> (16 August 2013).

Aidt, T.S., 2010. 'Corruption and Sustainable Development', Cambridge Working Papers in Economics 1061, Cambridge: University of Cambridge.

Alemayehu G., Weeks, J., Zerfu, D. & Degefa, J., 2004. Source of Growth in Ethiopia: A Background Paper for Government of Ethiopia. MOFED, Addis Ababa.

Alvarez, G.D., Tran, M. & Raina, R., 2011. 'MDGS: Sub-Saharan Africa: Overcoming Data Gaps and Ranking Progress', Fordham University, Department of Economics Discussion Paper Series no. 2011-01.

'Angola has a population of 24.3 million people', 2014. Macauhub, 16 October 2014 <http://www.macauhub.com.mo/en/2014/10/16/angola-has-a-population-of-24-3-million-people/> (13 January 2016).

Beegle, K., De Weerdt, J., Friedman, J., & Gibson, J., 2012. 'Methods of household consumption measurement through surveys: Experimental results from Tanzania.' Journal of Development Economics, 98.1: 3-18.

Black, R. & White, H., 2003. Targeting Development: critical perspectives on the Millennium Development Goals. Routledge.

Basu, K., 2014. 'Playing with and Understanding Purchasing Power Parities', World Bank's Chief Economist Blog, 7 May 2014 <http://blogs.worldbank.org/developmenttalk/playing-and-understanding-purchasing-power-parities> (1 March 2016).

Birdsall, N., Savedoff, W. D., Mahgoub, A., & Vyborny, K., 2010. Cash on delivery: a new approach to foreign aid. CGD Books.

Black, Richard, and Howard White, 2004. Targeting Development: Critical Perspectives on the Millennium Development Goals. Routledge: London.

Blades, Derek, 1980. 'What Do We Know about Levels and Growth ofOutput in Developing Countries? A Critical Analysis with Special Reference to Africa," in Economic Growth and Resources: Proceedings of the Fifth World Congress, International Economic Association,Tokyo , vol. 2, Trends and Factors , edited by R. C. O. Mathews(New York: St. Martin's Press, 1980), 60–77.

58. Ibid., 70.

Boden, Rebecca, and Debbie Epstein, 2006. "Managing the research imagination? Globalisation and research in higher education." Globalisation, Societies and Education 4.2: 223-36.

Broome, Andre and Joel Quirk, 2015. 'Governing the World at a Distance: The Practice of Global Benchmarking', *Review of International Studies* 41 (5), pp. 819-41.

Caldwell, J. C., & Okonjo, C., 1968. 'The population of tropical Africa.' Studies in Family Planning 1.29: 10-2.

Campbell, D.T., 1976. Assessing the Impact of Planned Social Change. The Public Affairs Center. Dartmouth College: Hanover.

Chandry, L., 2013. 'Counting the poor: Methods, problems and solutions behind the $1.25 a day global poverty estimates', Investments to End Poverty Working Paper 2013 May, Brookings Institute <http://reliefweb.int/sites/reliefweb.int/files/resources/Counting-the-poor.pdf> (1 March 2016).

Collins, M., 2015. 'Why Arbitrary Measures of Money-Laundering Risk are Nonsensical and Unfair', Center for Global Development <http://www.cgdev.org/blog/why-arbitrary-measures-money-laundering-risk-are-nonsensical-and-unfair> (1 March 2016).

Cooley, A. & Snyder, J., 2015. Ranking the World: Grading States as a Tool of Global Governance. Cambridge University Press: New York.

Cooper, F., 2002. Africa since 1940: The past of the present. Cambridge University Press: New York.

Coyle, D., 2014. GDP: A brief but affectionate history. Princeton University Press, Princeton.

Deaton, Angus, 2015. "Statistical Objectivity is a Cloak Spun from Political Yarn." Financial Times, 2 Nov 2015.

———, 2010. 'Price indexes, inequality, and the measurement of world poverty.' The American Economic Review 100.1: i-34.

Deaton, A. & Heston, A., 2010. 'Understanding PPPs and PPP-based national accounts', American Economic Journal: Macroeconomics, 2.4: 1-35.

'Décret : les résultats du recensement général de la population guinéenne publiés', 2015. Guineenews.org, 31 December 2015 < http://guineenews.org/decret-les-resultats-du-recensement-general-de-la-population-guineenne-publies/> (13 January 2016).

Deverajan, S., 2013. 'Africa's Statistical Tragedy', Review of Income and Wealth, 59.S1: S9–S15.

Durlauf, S.N. & Johnson, P.A. & Temple, J. R.W., 2005. 'Growth Econometrics', Handbook of Economic Growth, ed. 1, vol. 1, chap. 8, pp. 555-677. North Holland: Amsterdam.

Dyksta, S., Kenny, C., & Sandefur, J., 2014. 'Global Absolute Poverty Fell by Almost Half on Tuesday', Center for Global Development Blog, 2 May 2014 <http://www.cgdev.org/blog/global-absolute-poverty-fell-almost-half-tuesday> (1 March 2016).

Easterly, W., 2009. 'The burden of proof should be on interventionists—doubt is a superb reason for inaction', Boston Review, July/August 2009 < http://new.bostonreview.net/BR34.4/easterly.php> (15 January 2016).

———, 2003. 'Can Foreign Aid Buy Growth?', Journal of Economic Perspectives, 17.3: 23-48.

Feinstein, C.H., 2002. Making history count: a primer in quantitative methods for historians. Cambridge University Press, Cambridge; New York.

Fioramonti, L., 2013. Gross domestic problem. Zed Books, London.

Frankema, Ewout and Morten Jerven. 2014. 'Writing History Backwards or Sideways: Towards a Consensus on African Population, 1850-present', Economic History Review, Volume 67, Issue 4, pages 907–931.

Fukuda-Parr, S. and Yamin, A.E., 2013. 'The Power of Numbers: A critical review of MDG targets for human development and human rights', Development, 56.1: 58-65.

Geda, A. and Yimer, A., 2014. 'Growth, poverty and inequality in Ethiopia, 2000-2013: A macroeconomic appraisal', <https://www.researchgate.net/profile/Alemayehu_Geda/publication/264004552_Growth_Poverty_and_Inequality_in_Ethiopia_2000-2013_A_Macroeconomic_Appraisal/links/00b7d53c88d4ed0f9e000000.pdf> (15 January 2016).

Gibson, J., Beegle, K., De Weerdt, J. and Friedman, J., 2015. 'What does Variation in Survey Design Reveal about the Nature of Measurement Errors in Household Consumption?' Oxford Bulletin of Economics and Statistics, 77: 466–74.

Gisselquist, Rachel M., 2014. 'Developing and evaluating governance indexes: 10 questions', *Policy Studies*, 35.1: 513-531.

Guyer, Jane. 2014. *Marginal gains : monetary transactions in Atlantic Africa*, Chicago : The University of Chicago Press.

Hamermesh, D.S., 2013. 'Six Decades of Top Economics Publishing: Who and How', Journal of Economics Literature, 51.1: 162-72.

Herbst, J.I., 2000. States and power in Africa: Comparative lessons in authority and control. Princeton University Press: Princeton.

Hirschman, A.O., 1970. 'The Search for Paradigms as a Hindrance to Understanding', World Politics, 22.3: 329-43.

'How to lie with indices', The Economist, 8 November 2014: 17.

Independent Expert Advisory Group on a Data Revolution for Sustainable Development, 2014. A World That Counts: Mobilizing the data revolution for sustainable development < http://www.undatarevolution.org/report/> (11 Nov 2015).

International Monetary Fund, 2008. Review of Data Provision to the Fund for Surveillance Purposes, SM/08/76, 17 March 2008 (Washington, DC), <https://www.imf.org/external/np/pp/eng/2008/031708.pdf> (1 March 2016).

———, 2012a. Review of Data Provision to the Fund for Surveillance Purposes, SM/12/229, 28 August 2012 (Washington), <http://www.imf.org/external/np/pp/eng/2012/082812a.pdf> (1 March 2016)

———, 2012b. Ethiopia: 2012 Article IV Consultation, <http://www.imf.org/external/pubs/ft/scr/2012/cr12287.pdf> (15 January 2016).

———, 2013. Ethiopia: 2013 Article IV Consultation, <http://www.imf.org/external/pubs/ft/scr/2013/cr13308.pdf> (15 January 2016).

International Monetary Fund Evaluation Office, 2015. IMF Forecasts: Process, quality and country perspectives. IMF, Washington DC.

Kharas, H. & Chandy, L., 2014. 'Statistical Earthquakes', World Bank Future Development Blog, 7 May 2014 <http://blogs.worldbank.org/futuredevelopment/statistical-earthquakes> (1 March 2016).

Kravis, I.B., Heston, A.W. & Summers, R., 1978. 'Real GDP Per Capita for More Than One Hundred Countries', Economic Journal, 88.350: 215–42.

Jerven, Morten, 2010a. 'Random Growth in Africa? Lessons from an Evaluation of the Growth Evidence on Botswana, Kenya, Tanzania and Zambia', 1965–1995, Journal of Development Studies, 46:2, pp, 274 – 294.

———, 2010b. 'Accounting for the African Growth Miracle: The official Evidence, Botswana 1965-1995', Journal of Southern African Studies, 36:1, pp. 73-94,

———, 2010c. 'The Relativity of Poverty and Income: How Reliable Are African Economic Statistics?', African Affairs, 109:434, pp. 77-96, 2010.

———, 2011a. 'Revisiting the Consensus on Kenyan Economic Growth', 1964-1995, Journal of Eastern African Studies, 5:1, pp. 2-23.

———, 2011b. 'Growth, Stagnation or Retrogression? On the Accuracy of Economic Observations, Tanzania, 1961-2001', Journal of African Economies, 20:3, pp. 377-394.

———, 2011c. 'Users and Producers of African Income: Measuring the progress of African Economies', African Affairs, 110.439: 169-90.

———, 2012. 'An Unlevel Playing Field: National Income Estimates and Reciprocal Comparison in Global Economic History', Journal of Global History, 7: 1, pp. 107–128.

———, 2013a. Poor Numbers: How We Are Misled by African Development Statistics and What to Do about It. Cornell University Press, Cornell.

———, 2013b. 'For Richer, For Poorer: GDP Revisions and Africa's Statistical Tragedy', African Affairs, 112.446: 138-47.

———, 2014a. Economic Growth and Measurement Reconsidered in Botswana, Kenya, Tanzania, and Zambia, 1965-1995. Oxford University Press: Oxford.

———, 2014b. 'What Does Nigeria's new GDP number actually mean?', African Arguments, 8 April 2014 <http://africanarguments.org/2014/04/08/what-does nigerias-new-gdp-number-actually-mean-by-morten-jerven/> [December 2015].

———, 2014c. 'Benefits and costs of the data for development targets for the Post-2015 Development Agenda.' Data for Development Assessment Paper Working Paper, September. Copenhagen: Copenhagen Consensus Center.

———, 2014d. 'Benefits and Costs of the Data for Development Targets for the Post-2015 Development Agenda', Post-2015 Consensus Assessment, Copenhagen: Copenhagen Consensus Center.

———, 2014e. 'Poor numbers and what to do about them', The Lancet, 383.9917: 594-5.

———, 2014f. 'The Political Economy of Agricultural Statistics and Input Subsidies: Evidence from India, Nigeria and Malawi', Journal of Agrarian Change, 14.1: 129-45.

———, 2014g. 'Why saying 'seven out of ten fastest growing economies in Africa' carries no real meaning', African Arguments, 26 August 2014 <http://africanarguments.org/2014/08/26/why-saying-seven-out-of-ten-fastest-

growing-economies-are-in-africa-carries-no-real-meaning-by-morten-jerven/ > (1 March 2016).

———, 2015a. Measuring African Development Past and Present. Routledge: Abingdon.

———, 2015b. Africa: Why Economists Get It Wrong. Zed Books: London.

Jerven, M. & Johnston, D., 2015. Statistical tragedy in Africa? Evaluating the database for African economic development. Routledge: Abingdon.

Jolliffe, D., 2001. 'Measuring absolute and relative poverty: the sensitivity of estimated household consumption to survey design', Journal of Economic and Social Measurement, 27.1: 1-23.

K'akumu, O.A., 2007. 'Construction statistics review for Kenya', Construction Management and Economics, 25.3: 315-26.

Kalungi, N., 2013. 'Minister blames insufficient statistics for slowed development', Daily Monitor, 23 January 2013 <http://www.monitor.co.ug/Business/Minister-blames-insufficient-statistics-for-slowed-development/-/688322/1672422/-/9v1bulz/-/index.html> (13 January 2016).

Kerner, A., Jerven, M., & Beatty, A., 2015. 'Does it pay to be poor? Testing for systematically underreported GNI estimates', The Review of International Organizations: 1-38.

Kilic, T., Sohnesen, T., 2015. 'Same question but different answer: experimental evidence on questionnaire design's impact on poverty measured by proxies.' Policy Research working paper; no. WPS 7182. Washington, DC: World Bank Group.

Leamer, E., 1983. 'Let's Take the CON Out of Econometrics', American Economic Review, 73.1: 31-43.

Manning, Richard. 2009. 'Using Indicators to Encourage Development: Lessons from the Millennium Development Goals', Danish Institute for International Studies (DIIS), Report 2009.1: 1-100.

Mikkelsen, L, Phillips, D.E., AbouZhar, C., Setel, P.W., de Savigny, D., Lozano, R., & Lopez, A.D., 'A global assessment of civil registration and vital statistics systems: monitoring data quality and progress', The Lancet, 386.10001: 1395-406.

Musoke, R., 2014. 'Shs 68 trillion economy', The Independent, 15 December 2014 <http://www.independent.co.ug/business/business-news/9580-shs-68-trillion-economy-> (13 January 2016).

Olukya, G., 2014. 'Uganda's Population Registers a Massive Surge', The Africa Report, 19 November 2014 <http://www.theafricareport.com/East-Horn-Africa/ugandas-population-registers-a-massive-surge.html> (13 January 2016).

Pastor, G., 2009. 'A Framework to Assess the Effectiveness of IMF Technical Assistance in National Accounts', IMF Working Paper WP/09/269 <https://www.imf.org/external/pubs/ft/wp/2009/wp09269.pdf> (1 March 2016).

'Population: We happy few', 2015. The Economist, 20 June 2015 <http://www.economist.com/news/special-report/21654359-nigerias-population-has-been-systematically-exaggerated-we-happy-few> (13 January 2016).

Porter, T.M., 1995. Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton University Press: Princeton.

Randall, Sara and Ernestina Coast, 2015. 'Poverty in African Households: the Limits of Survey and Census Representations', The Journal of Development Studies, 51:2, 162-177.

Sanga, D., 2-11. 'The Challenges of Monitoring and Reporting on the Millennium Development Goals in Africa by 2015 and Beyond', African Statistical Journal, 12: 104-18.

Sachs, J., 2009. 'Homegrown Aid', The New York Times, 8 April 2009: A23.

Sandefur, J. & Glassman, A., 2015. 'The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics', The Journal of Development Studies, 51.2: 116-32.

Scidevnet (2014), 'Frustration' over old or missing data in MDGs report', SciDevNet, 8 July 2014 <http://www.scidev.net/global/mdgs/news/frustration-over-old-or-missing-data-in-mdgs-report.html> (1 March 2016).

Seers, D., 1952-1953. 'The Role of National Income Estimates in the Statistical Policy of an Under-developed Area', The Review of Economic Studies, 20.3: 159-68.

Serajuddin, U., Uematsu, H., Wieser, C., Yoshida, N., & Dabalen, A.L., 2015. 'Data deprivation: another deprivation to end', World Bank Policy Research Working Paper no. WPS 7252 <http://documents.worldbank.org/curated/en/2015/04/24426058/data-deprivation-another-deprivation-end> (1 March 2016).

Srinivasan, T.N., 1994. 'The Data Base for Development Analysis: An Overview', Journal of Development Economics, 44.1: 3-27.

Stern, N., 1989. 'The Economics of Development: A Survey', The Economic Journal, 99.397: 597–685.

Stiglitz, J.E., Sen, A., and Fitoussi, J.-P., 2010. Mismeasuring our lives: Why GDP doesn't add up. New Press: New York.

Subramanian, S., 2012. 'The Poverty Line', Oxford India Short Introductions. Oxford University Press.

Summers, R. & Heston, A., 1991. 'The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–1988.' Quarterly Journal of Economics, 106.2: 327–68.

Taylor, A.M. & Taylor, M.P., 2004. 'The Purchasing Power Parity Debate', Journal of Economic Perspectives, 18.4: 135-58.

United Nations High Level Panel, 2013. "A new global partnership: Eradicate poverty and transform economies through sustainable development." Final Report of the UN High-Level Panel of Eminent Persons on the Post-2015 Development Agenda.

United Nations Statistics Division, 2016. Millennium Development Goals Indicators, < http://unstats.un.org/unsd/mdg/SeriesDetail.aspx?srid=699> (21 January 2016).

Vandemoortele, J., 2011. 'The MDG Story: Intention Denied', Development and Change, 42.1: 1-21.

Ward, M., 1971. 'Review of the Measurement of Real Product', The Economic Journal, 81.324: 974-77.

———, 2004. Quantifying the world: UN ideas and statistics. Indiana University

Press: Bloomington.

Welle, K., 2014. 'Monitoring performance or performing monitoring? Exploring the power and political dynamics underlying monitoring the MDG for rural water in Ethiopia', Canadian

Journal of Development Studies / Revue canadienne d'études du développement, 35:1: 155-69.

World Development Indicators, 2015. World Bank, <http://data.worldbank.org/data-catalog/world-development-indicators> (13 January 2016).

Appendix 1 - EIU--Country Reports: National Accounts' Data's Timeliness & Information Source (source: Pastor, 2009, p. 18)

| | Date of EIU Report | Latest "actual" GDP data | "Estimated GDP data" (source) | Other referred sources on national accounts data |
|---|---|---|---|---|
| Benin | April 2009 | 2005 | 2006-08, EIU | Association interprofessionelle du coton; IMF |
| Burkina Faso | May 2009 | 2006 | 2007-08, EIU | IMF |
| Côte d'Ivoire | May 2009 | 2007 | 2008-10, EIU | WAEMU; Direction Générale des Douanes; Direction de la Conjoncture et de la Prévision économique; IMF |
| Guinea | June 2009 | 2005 | 2006-08, EIU | IMF |
| Guinea-Bissau | April 2009 | 2006 | 2007-08, EIU | IMF |
| Mali | May 2009 | 2006 | 2007-08, EIU | BCEAO; Ministry of Agriculture; IMF |
| Mauritania | April 2009 | 2006 | 2007-08, EIU | IMF |
| Niger | May 2009 | 2006 | 2007-08, EIU, IMF | IMF |
| Senegal | May 2009 | 2008 | 2009-10, EIU | Agence Nationale de la Statistique et de la Démographie, Ministry of Economy and Financy; IMF |
| Togo | April 2009 | 2006 | 2007-08, EIU | BCEAO; Ministry of Agriculture; IMF |
| Botswana | May 2009 | 2008 | 2008-10, EIU | Central Statistics Office; Bank of Botswana; Botswana Financial Statistics; Annual Report; IMF |
| The Gambia | April 2009 | 2007 | 2008, EIU | IMF |
| Kenya | May 2009 | 2007 | 2008-10, EIU | Central Bank of Kenya; IMF |
| Mauritius | May 2009 | 2008 | 2009-10, EIU | Central Statistics Office; Ministry of Tourism, Leisure and External Communications; Ministry of Finance and Economic Development, IMF |
| Mozambique | May 2009 | 2007 | 2008-10, EIU | UN Food and Agriculture Organization; IMF |
| Namibia | May 2009 | 2008 | 2009-10, EIU | Bank of Namibia; Central Bureau of Statistics; Irwin, Jacobs, Green/Institute for Public Policy Research, Windhoek; IMF |
| South Africa | May 2009 | 2007 | 2008-10, EIU | Statistics South Africa, South African Reserve Bank; IMF |
| Tanzania | May 2009 | 2007 | 2008-10, EIU | Barrick Gold; UN Food and Agriculture Organization; Bank of Tanzania; IMF |
| Uganda | May 2009 | 2007 | 2008-10, EIU | Bank of Uganda; Uganda Bureau of Statistics; IMF |
| Zambia | June 2009 | 2007 | 2008-10, EIU | Ministry of Agriculture and Cooperatives, Bank of Zambia; IMF |

Appendix 2 – Relative Reliability of Estimates in PWT 6.1

| | |
|---|---|
| C | Armenia, Bangladesh, Benin, Bolivia, Burkina Faso, Burundi, Cameroon, Republic of the Congo, Côte d'Ivoire, Egypt, El Salvador, Ethiopia, Gambia, Georgia, Ghana, Guatemala, Guinea, Honduras, India, Indonesia, Kenya, |

| | |
|---|---|
| | Kyrgyz Republic, Madagascar, Malawi, Mali, Mauritania, Moldova, Morocco, Nepal, Nicaragua, Nigeria, Pakistan, Philippines, Rwanda, Senegal, Sierra Leone, Sri Lanka, Swaziland, Syrian Arab Republic, Tanzania, Ukraine, Vietnam, Zambia, Zimbabwe |
| D | Bhutan, Cabo Verde, Cambodia, Central African Republic, Chad, Comoros, Democratic Republic of the Congo, Djibouti, Eritrea, Guinea-Bissau, Guyana, Haiti, Lao PDR, Lesotho, Liberia, Mozambique, Myanmar, Niger, Papua New Guinea, São Tomé and Principe, Somalia, Sudan, Tajikistan, Togo, Uganda, Uzbekistan, Yemen |
| Not Available | Afghanistan, Kiribati, North Korea, Kosovo, Micronesia, Samoa, Solomon Islands, South Sudan, Timor-Leste, Vanuatu, West Bank and Gaza |

Appendix 3 – Average Statistical Capacity Score, 2010-2015

| | |
|---|---|
| 80+ | Armenia, Egypt, Arab Rep., El Salvador, Georgia, Indonesia, Kyrgyz Republic, Moldova, Philippines, Ukraine |
| 60-80 | Bangladesh, Benin, Bhutan, Bolivia, Burkina Faso, Cabo Verde, Cambodia, Cameroon, Chad, Ethiopia, Gambia, Ghana, Guatemala, Honduras, India, Lao PDR, Lesotho, Madagascar, Malawi, Mali, Mauritania, Morocco, Mozambique, Nepal, Nicaragua, Niger, Nigeria, Pakistan, Rwanda, São Tomé and Principe, Senegal, Sri Lanka, Swaziland, Tajikistan, Tanzania, Uganda, Vietnam, West Bank and Gaza |
| <60 | Afghanistan, Burundi, Central African Republic, Comoros, Democratic Republic of the Congo, Republic of the Congo, Côte d'Ivoire, Djibouti, Eritrea, Guinea, Guinea-Bissau, Guyana, Haiti, Kenya, Kiribati, Kosovo, Liberia, Micronesia, Myanmar, Papua New Guinea, Samoa, Sierra Leone, Solomon Islands, Somalia, South Sudan, Sudan, Syrian Arab Republic, Timor-Leste, Togo, Uzbekistan, Vanuatu, Yemen, Zambia, Zimbabwe |
| Not Available | North Korea |

Appendix 4 – Frequency of GDP Observations in the International Financial Statistics Records

| | |
|---|---|
| >75% | Bangladesh, Benin, Bolivia, Burkina Faso, Burundi, Cambodia, Cameroon, Côte d'Ivoire, Egypt, El Salvador, Georgia, Guatemala, Guinea-Bissau, Haiti, Honduras, India, Indonesia, Kenya, Kyrgyz Republic, Lao PDR, Madagascar, Malawi, Mali, Morocco, Mozambique, Nepal, Nicaragua, Niger, Pakistan, Philippines, Rwanda, Senegal, Sierra Leone, South Sudan, Sri Lanka, Timor-Leste, Togo, Uganda, Ukraine, Vanuatu, Vietnam, Yemen |
| 40%-75% | Chad, Gambia, Lesotho, Mauritania, Nigeria, Swaziland, Syrian Arab Republic, Zambia |
| 0% - <40% | Democratic Republic of the Congo, Solomon Islands, Myanmar, Cabo Verde, Guyana, Samoa, Zimbabwe, Central African Republic, Papua New Guinea, Swaziland |
| 0% | Afghanistan, Armenia, Bhutan, Comoros, Republic of the Congo, Djibouti, Eritrea, Ethiopia, Ghana, Guinea, Kiribati, North Korea, Kosovo, Liberia, Micronesia, Moldova, São Tomé and Principe, Somalia, Sudan, Tajikistan, Tanzania, Uzbekistan, West Bank and Gaza |

Appendix 5 – Poverty Data Deprivation in the Developing World

| | |
|---|---|
| Extreme | Eritrea, Guinea-Bissau, Guyana, Haiti, Kiribati, North Korea, Micronesia, Myanmar, Papua New Guinea, Samoa, Solomon Islands, Somalia, Vanuatu, Zimbabwe |
| Moderate | Afghanistan, Burundi, Cabo Verde, Cameroon, Chad, Comoros, Democratic Republic of the Congo, Côte d'Ivoire, Djibouti, The Gambia, Ghana, Kenya, Lesotho, Liberia, Morocco, Mozambique, São Tomé and Principe, South Sudan, Sudan, Swaziland, Syrian Arab Republic, Timor-Leste, Uzbekistan, Yemen |
| Vulnerable | Benin, Burkina Faso, Republic of the Congo, Ethiopia, Malawi, Nepal, Nigeria, Rwanda, Senegal, Sierra Leone |
| Minimum | Bangladesh, Central African Republic, Egypt, India, Lao PDR, Madagascar, Mali, Mauritania, Tanzania, Togo |
| Satisfactory | Armenia, Bhutan, Bolivia, Cambodia, El Salvador, Georgia, Guatemala, Guinea, Honduras, Indonesia, Kosovo, Kyrgyz Republic, Moldova, Nicaragua, Niger, Pakistan, Philippines, Sri Lanka, Tajikistan, Uganda, Ukraine, Vietnam, West Bank and Gaza, Zambia |

Appendix 6 – Participation in ICP Rounds

| | |
|---|---|
| 5+ | Cameroon, Côte d'Ivoire, Ethiopia, India, Kenya, Madagascar, Malawi, Mali, Morocco, Nigeria, Pakistan, Philippines, Senegal, Sri Lanka, Tanzania, Zambia |
| 3-4 | Armenia, Bangladesh, Benin, Bhutan, Bolivia, Burkina Faso, Burundi, Cabo Verde, Central African Republic, Chad, Republic of the Congo, Egypt, Gambia, Ghana, Guinea-Bissau, Indonesia, Lao PDR, Lesotho, Mauritania, Mozambique, Nepal, Niger, Rwanda, Sierra Leone, Sudan, Swaziland, Syrian Arab Republic, Togo, Uganda, Ukraine, Yemen, Zimbabwe |
| 0-2 | Afghanistan, Cambodia, Comoros, Democratic Republic of the Congo, Djibouti, El Salvador, Eritrea, Georgia, Guatemala, Guinea, Guyana, Haiti, Honduras, Kiribati, North Korea, Kosovo, Kyrgyz Republic, Liberia, Micronesia, Moldova, Myanmar, Nicaragua, Papua New Guinea, Samoa, São Tomé and Principe, Solomon Islands, Somalia, South Sudan, Tajikistan, Timor-Leste, Uzbekistan, Vanuatu, Vietnam, West Bank and Gaza |

Appendix 7 – Participation in Last Three Census Rounds

| | |
|---|---|
| 3 Rounds | Armenia, Bangladesh, Benin, Bolivia, Burkina Faso, Cabo Verde, Republic of the Congo, Côte d'Ivoire, Egypt, Gambia, Georgia, Guyana, Honduras, India, Indonesia, Kenya, Kiribati, Kyrgyz Republic, Lao PDR, Lesotho, Malawi, Mali, Mauritania, Micronesia, Moldova, Morocco, Nepal, Niger, Papua New Guinea, Philippines, Rwanda, Samoa, São Tomé and Principe, Senegal, Solomon Islands, Swaziland, Tajikistan, Tanzania, Timor-Leste, Uganda, Vanuatu, Vietnam, Zambia, Zimbabwe |
| 2 Rounds | Burundi, Cambodia, Cameroon, Central African Republic, Chad, Comoros, El Salvador, Ethiopia, Ghana, Guatemala, Guinea, Guinea-Bissau, North Korea, Mozambique, Nicaragua, Nigeria, Sierra Leone, South Sudan, Sri Lanka, Sudan, Syrian Arab Republic, Ukraine |
| 1 Round | Bhutan, Djibouti, Haiti, Liberia, Madagascar, Myanmar, Pakistan, Somalia, Togo, Uzbekistan, Yemen\ |
| 0 Rounds | Afghanistan, Democratic Republic of the Congo, Eritrea, Kosovo, West |

| | |
|---|---|
| | Bank and Gaza |

Appendix 8 – Health and Education Metadata for Low-Income Countries (source: IMF GDDS) (Note: metadata self-reported)

| | Health Data | Education Data | Notes | Last Updated |
|---|---|---|---|---|
| Afghanistan | Not available | Not available | | |
| Benin | Not available | Not available | | |
| Burkina Faso | Compiled Annually | Compiled Annually | | 2001 |
| Burundi | Not available | Not available | | |
| Cambodia | Reported monthly, published annually | Collected Annually | | 2007 |
| Central African Republic | Collected monthly/quarterly; published annually | Compiled Annually | | 2004 |
| Chad | Collected annually, every 5 years, and every 10 years depending on source | Collected annually, every 5 years, and every 10 years depending on source | Data available from administrative sources, census, DHS, Consumption and Informal Sectory Survey | 2002 |
| Comoros | Not available | Not available | | |
| Congo, Dem. Rep. | Published annually | Published annually | Monthly reports recommended by IMF staff | 2004 |
| Eritrea | Not included in GDDS | | | |
| Ethiopia | Collected annually | Collected annually | DHS Planned for every 5 years | 2003 / 2006 |
| Gambia, The | Collected monthly at village/basic health facilities, published sporadically | Collected annually | | 2003 |
| Guinea | Not available | Prepared annually | | 2002 |
| Guinea-Bissau | Not dissimenated by national agency; WHO disseminates quarterly and annually | Infrequently | Education data is supposed to be published annually by national ministry but last publication was 1993/94 | 2001 |
| Haiti | Not available | Infrequently | Intended to be annual, financing constraints have made this impossible though | 2009 |
| Korea, Dem. Rep. | Not included in GDDS | | | |
| Liberia | Produced annually | Annually, with gaps | Supposed to be annual, financial | 2011 / 2013 |

| | | | | |
|---|---|---|---|---|
| | | | constraint has led to lags and gaps | |
| Madagascar | Quarterly at district/provincial level, annually at national level | Annual reports with supplemental surveys every 3 years | | 2004 |
| Malawi | Infrequently, via DHS | Not available | | 2007 |
| Mali | Published annually | Compiled Annually | | 2001 |
| Mozambique | Collected regularly | Collected every 4 months | Some lag in health data due to remote regions | 2009 |
| Nepal | Compiled monthly, quarterly, annually | Annually | | 2011 |
| Niger | Varies, mostly annual with some monthly reporting | Not available | | 2001 |
| Rwanda | Weekly, monthly, quarterly, annual reporting | Annually | | 2011 / 2009 |
| Sierra Leone | Mostly annual, some monthly reporting | Annual when funds available | No education data since 2001 due to funding issues | 2006 |
| Somalia | Not included in GDDS | | | |
| South Sudan | Not included in GDDS | | | |
| Tanzania | Monthly, quarterly, annually | Annually | | 2014 |
| Togo | Complied daily/weekly (health facilities, monthly | Not available | Understaffing results in longer-than-established wait times for publication of results | 2002 |
| Uganda | Periodic, and annual | Annual | Variety of health surveys included in GDDS report | 2013 / 2002 |
| Zimbabwe | Compiled Monthly | Compiled Annually | | 2013 |

## Appendix 9 – Sandefur and Glassman's (2015, p. 125) Changes in Primary School Net Enrolment

| Country | Start | End | Admin. Data Enrolment Change | Survey Data Enrolment Change | Gap |
|---|---|---|---|---|---|
| Kenya | 1998 | 2003 | 17.8% | -3.6% | 21.4% |
| Rwanda | 2005 | 2010 | 16.9% | 1.9% | 15.0% |
| Ethiopia | 2000 | 2005 | 21.6% | 12.0% | 9.6% |
| Cameroon | 1991 | 2011 | 23.7% | 14.6% | 9.1% |
| Burkina Faso | 1993 | 1999 | 6.4% | -1.6% | 8.0% |
| Kenya | 2003 | 2008 | 7.8% | 0.0% | 7.8% |

| | | | | | |
|---|---|---|---|---|---|
| Benin | 1996 | 2006 | 25.1% | 18.3% | 6.8% |
| Burkina Faso | 2003 | 2010 | 21.5% | 16.4% | 5.1% |
| Eritrea | 1995 | 2002 | 16.7% | 13.7% | 3.0% |
| Niger | 1992 | 2006 | 20.9% | 18.5% | 2.4% |
| Ethiopia | 2005 | 2011 | 24.6% | 22.3% | 2.3% |
| Guinea | 1999 | 2005 | 25.1% | 23.4% | 1.7% |
| Senegal | 2005 | 2010 | 3.3% | 2.3% | 1.0% |
| Namibia | 1992 | 2000 | 5.5% | 4.8% | 0.7% |
| Burkina Faso | 1999 | 2003 | 3.2% | 3.0% | 0.2% |
| Tanzania | 1999 | 2004 | 36.9% | 38.1% | -1.2% |
| Tanzania | 1992 | 1996 | -1.9% | 1.1% | -3.0% |
| Nigeria | 1999 | 2003 | 4.3% | 7.5% | -3.2% |
| Nigeria | 2003 | 2008 | -6.8% | -2.2% | -4.6% |
| Tanzania | 1996 | 1999 | 0.6% | 7.7% | -7.1% |
| Lesotho | 2004 | 2009 | -2.0% | 7.9% | -9.9% |

Appendix 10 – AIV: Adequacy for Surveillance rating of the country's statistics in 2012, or latest available prior to 2012.

| Income Classification | AIV Score | Countries |
|---|---|---|
| Low Income | A | Kyrgyz R., Uganda |
| | B | Afghanistan, Albania, Angola, Armenia, Azerbaijan, Bangladesh, Benin, Bhutan, Bolivia, Burkina Faso, Cambodia, Cameroon, Cape Verde, CAR, Chad, Côte d'Ivoire, Djibouti, Dominica, DRC, Ethiopia, The Gambia, Georgia, Ghana, Grenada, Guinea, Guyana, Honduras, India, Kenya, Kiribati, Lesotho, Madagascar, Malawi, Maldives, Mali, Mauritania, Moldova, Mongolia, Mozambique, Nepal, Nicaragua, Niger, Nigeria, Papua New Guinea, Pakistan, Republic of Congo, Rwanda, Sao Tome and Principe, Samoa, Senegal, Solomon Islands, Sri Lanka, St. Lucia, St. Vincent and The Grenadines, Sudan, Tajikistan, Tanzania, Togo, Tonga, Vanuatu, Vietnam, Yemen, Zambia, Zimbabwe |
| | C | Burundi, Comoros, Eritrea, Guinea-Bissau, Haiti, Lao P.D.R., Liberia, Myanmar, Sierra Leone, Timor-Leste, Uzbekistan |
| | n/a | Somalia |
| Non Low-Income | A | Finland, Germany, Hungary, Ireland, Mexico, Morocco, New Zealand, Poland, Portugal, Slovenia, South Africa, South Korea, Spain, Sweden, USA |
| | B | Algeria, Antigua, Argentina, Botswana, China, Costa Rica, Egypt, Fiji, Indonesia, Israel, Kazakhstan, Malaysia, Namibia, Paraguay, Saudi Arabia, Thailand, Turkey, Ukraine, Uruguay |
| | C | Equatorial Guinea, Greece, Jamaica |

Appendix 11 – Availability of Datasets, by country*

| | Availability | | | | |
|---|---|---|---|---|---|
| | GDP | Pov. | PPP | Pop. | MDG/ SDG** |
| Afghanistan | 0.00 | 0.25 | 0.00 | 0.00 | 0.00 |
| Armenia | 0.00 | 1.00 | 0.67 | 1.00 | SDDS |
| Bangladesh | 1.00 | 0.75 | 0.67 | 1.00 | 1.00 |
| Benin | 1.00 | 0.50 | 0.67 | 1.00 | 0.00 |
| Bhutan | 0.00 | 1.00 | 0.67 | 0.33 | 1.00 |
| Bolivia | 1.00 | 1.00 | 0.67 | 1.00 | 1.00 |
| Burkina Faso | 0.93 | 0.50 | 0.67 | 1.00 | 1.00 |
| Burundi | 0.93 | 0.25 | 0.67 | 0.67 | 0.00 |
| Cabo Verde | 0.29 | 0.25 | 0.67 | 1.00 | 0.00 |
| Cambodia | 0.93 | 1.00 | 0.33 | 0.67 | 1.00 |
| Cameroon | 0.93 | 0.25 | 1.00 | 0.67 | 0.50 |
| Central African Republic | 0.36 | 0.75 | 0.67 | 0.67 | 1.00 |
| Chad | 0.64 | 0.25 | 0.67 | 0.67 | 1.00 |
| Comoros | 0.00 | 0.25 | 0.33 | 0.67 | 0.00 |
| Congo, Dem. Rep. | 0.07 | 0.25 | 0.33 | 0.00 | 0.00 |
| Congo, Rep. | 0.00 | 0.50 | 0.67 | 1.00 | 0.75 |
| Côte d'Ivoire | 0.93 | 0.25 | 1.00 | 1.00 | 1.00 |
| Djibouti | 0.00 | 0.25 | 0.33 | 0.33 | 0.00 |
| Egypt, Arab Rep. | 1.00 | 0.75 | 0.67 | 1.00 | SDDS |
| El Salvador | 1.00 | 1.00 | 0.33 | 0.67 | SDDS |
| Eritrea | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Ethiopia | 0.00 | 0.50 | 1.00 | 0.67 | 1.00 |
| Gambia, The | 0.64 | 0.25 | 0.67 | 1.00 | 0.75 |
| Georgia | 1.00 | 1.00 | 0.33 | 1.00 | SDDS |
| Ghana | 0.00 | 0.25 | 0.67 | 0.67 | 0.75 |
| Guatemala | 1.00 | 1.00 | 0.33 | 0.67 | 0.00 |
| Guinea | 0.00 | 1.00 | 0.33 | 0.67 | 0.50 |
| Guinea-Bissau | 0.93 | 0.00 | 0.67 | 0.67 | 0.50 |
| Guyana | 0.29 | 0.00 | 0.00 | 1.00 | 0.50 |
| Haiti | 1.00 | 0.00 | 0.33 | 0.33 | 0.25 |
| Honduras | 0.93 | 1.00 | 0.33 | 1.00 | 0.00 |
| India | 1.00 | 0.75 | 1.00 | 1.00 | SDDS |
| Indonesia | 1.00 | 1.00 | 0.67 | 1.00 | SDDS |
| Kenya | 0.86 | 0.25 | 1.00 | 1.00 | 1.00 |
| Kiribati | 0.00 | 0.00 | 0.33 | 1.00 | 0.75 |
| Korea, Dem Rep. | 0.00 | 0.00 | 0.00 | 0.67 | 0.00 |
| Kosovo | 0.00 | 1.00 | 0.00 | 0.00 | 1.00 |
| Kyrgyz Republic | 1.00 | 1.00 | 0.33 | 1.00 | SDDS |
| Lao PDR | 0.93 | 0.75 | 0.67 | 1.00 | n/a |
| Lesotho | 0.57 | 0.25 | 0.67 | 1.00 | 0.50 |

| | | | | | |
|---|---|---|---|---|---|
| Liberia | 0.00 | 0.25 | 0.33 | 0.33 | 0.75 |
| Madagascar | 0.93 | 0.75 | 1.00 | 0.33 | 1.00 |
| Malawi | 0.93 | 0.50 | 1.00 | 1.00 | 0.25 |
| Mali | 0.93 | 0.75 | 1.00 | 1.00 | 1.00 |
| Mauritania | 0.50 | 0.75 | 0.67 | 1.00 | 1.00 |
| Micronesia, Fed. Sts. | 0.00 | 0.00 | 0.33 | 1.00 | 0.00 |
| Moldova | 0.00 | 1.00 | 0.33 | 1.00 | SDDS |
| Morocco | 0.93 | 0.25 | 1.00 | 1.00 | SDDS |
| Mozambique | 0.93 | 0.25 | 0.67 | 0.67 | 1.00 |
| Myanmar | 0.21 | 0.00 | 0.33 | 0.33 | 0.00 |
| Nepal | 1.00 | 0.50 | 0.67 | 1.00 | 1.00 |
| Nicaragua | 0.86 | 1.00 | 0.33 | 0.67 | 0.00 |
| Niger | 0.93 | 1.00 | 0.67 | 1.00 | 0.50 |
| Nigeria | 0.57 | 0.50 | 1.00 | 0.67 | 0.25 |
| Pakistan | 1.00 | 1.00 | 1.00 | 0.33 | 1.00 |
| Papua New Guinea | 0.36 | 0.00 | 0.33 | 1.00 | 0.00 |
| Philippines | 1.00 | 1.00 | 1.00 | 1.00 | SDDS |
| Rwanda | 0.86 | 0.50 | 0.67 | 1.00 | 1.00 |
| Samoa | 0.29 | 0.00 | 0.33 | 1.00 | 0.50 |
| São Tomé and Principe | 0.00 | 0.25 | 0.33 | 1.00 | 1.00 |
| Senegal | 0.93 | 0.50 | 1.00 | 1.00 | 1.00 |
| Sierra Leone | 0.93 | 0.50 | 0.67 | 0.67 | 0.75 |
| Solomon Islands | 0.14 | 0.00 | 0.33 | 1.00 | 0.00 |
| Somalia | 0.00 | 0.00 | 0.00 | 0.33 | 0.00 |
| South Sudan | 0.83 | 0.25 | 0.00 | 0.67 | 0.00 |
| Sri Lanka | 1.00 | 1.00 | 1.00 | 0.67 | SDDS |
| Sudan | 0.00 | 0.25 | 0.67 | 0.67 | 1.00 |
| Swaziland | 0.43 | 0.25 | 0.67 | 1.00 | 1.00 |
| Syrian Arab Republic | 0.71 | 0.25 | 0.67 | 0.67 | 0.00 |
| Tajikistan | 0.00 | 1.00 | 0.33 | 1.00 | 1.00 |
| Tanzania | 0.00 | 0.75 | 1.00 | 1.00 | 1.00 |
| Timor-Leste | 0.79 | 0.25 | 0.00 | 1.00 | 0.50 |
| Togo | 0.93 | 0.75 | 0.67 | 0.33 | 0.25 |
| Uganda | 0.93 | 1.00 | 0.67 | 1.00 | 1.00 |
| Ukraine | 0.86 | 1.00 | 0.67 | 0.67 | SDDS |
| Uzbekistan | 0.00 | 0.25 | 0.00 | 0.33 | n/a |
| Vanuatu | 0.93 | 0.00 | 0.33 | 1.00 | 0.50 |
| Vietnam | 0.86 | 1.00 | 0.33 | 1.00 | 1.00 |
| West Bank and Gaza | 0.00 | 1.00 | 0.00 | 0.00 | SDDS |
| Yemen, Rep. | 0.86 | 0.25 | 0.67 | 0.33 | 1.00 |
| Zambia | 0.64 | 1.00 | 1.00 | 1.00 | 0.50 |
| Zimbabwe | 0.29 | 0.00 | 0.67 | 1.00 | 1.00 |

*Methodologies used:

- GDP: The figure in this column is the result of dividing the number of annual GDP observations for the country in the IFS by a maximum possible score of 14 (2001-2014). A score of 1.0 reflects that a GDP statistic is available every year and a score of 0.0 reflects that none are available.
- Poverty: The ordinal variables related to data deprivation are coded as follows:
  - Extreme: 0.0
  - Moderate: 0.25
  - Vulnerable: 0.5
  - Minimum: 0.75
  - Satisfactory: 1.0
- PPP: Countries are placed into one of four groups depending on how many ICP rounds they have participated in. These were scored as follows:
  - 0 rounds: 0.0
  - 1-2 rounds: 0.33
  - 3-4 rounds: 0.67
  - 5+ rounds: 1.0
- Population: Countries are placed into one of four groups depending on how many of the last three census rounds they participated in. These were scored as follows:
  - 0 rounds: 0.0
  - 1 round: 0.33
  - 2 rounds: 0.67
  - 3 rounds: 1.0
- MDG/SDGs: Two dimensions are investigated – health and education. For each dimension, a score of 1.0 is given if the GDDS indicates annual reporting, a score of 0.0 if there is no information available, and a score of 0.5 if reporting is infrequent, less regular than annually, or subject to other issues. The score reflected in the average score of these two dimensions [ (health + education) / 2]
- **Scores of SDDS in this column reflect that the country is included in the IMF's Special Data Dissemination Standard, a separate database that runs parallel to the GDDS.

The scoring on reliability of the data is not done on country by country basis, but are given average values, as explained in the text.