EQUALITAS
Economics of Inequality and Poverty Analysis

EQUALITAS Working Paper No. 15

Human Development Index-like Small Area Estimates for Africa computed from IPUMS-International integrated census microdata

Iñaki Permanyer

Albert Esteve-Palos

Joan Garcia

Robert McCaa

2013

Title: Human Development Index-like Small Area Estimates for Africa computed from IPUMS-International integrated census microdata

Authors: Iñaki Permanyer[1], Albert Esteve-Palos[1], Joan Garcia[1], and Robert McCaa[2]

[1]Demographic Studies Center, Autonomous University of Barcelona

[2]University of Minnesota Population Center. Corresponding author: rmccaa@umn.edu

Africa is the continent that could benefit the least
from the [HDI] methodology presented in this paper
given the scarcity of census data in that region of the world.
--Permanyer, World Development, 2013

1. Introduction

Is the greater "statistical tragedy" in Africa (Devarajan 2013) the scarcity of census data, or the lack of access to the existing data? Is the problem "Poor Numbers" (Jerven 2013) or inaccessible numbers? In each decade since the 1970s, at least 80% of the continent's population was censused, yet much of the microdata are not available for scientific or policy research. From the late 1980s, twenty five countries entrusted microdata to the African Census Analysis Project (ACAP), amassing a stock of microdata for 47 censuses. Nonetheless, in recent years, the project seems moribund with no published research, nor cosmetic touch-ups to the ACAP website since 2007. The director, Dr. Tukufu Zuberi, no longer permits access, even to researchers wishing to study their own country nor does he respond to requests to repatriate copies to the official statistical office-owner of the microdata.

In 1999, the Minnesota Population center began a global initiative, IPUMS-International, offering free, internet access (www.ipum.org/international) to integrated census microdata for researchers world-wide under a single license agreement with National Statistical Office partners. Microdata for 69 countries, totaling 480 million person records (212 samples), are accessible for research. The June 2013 release will increase the number of countries to 74 with 238 samples and over 540 million person records. IPUMS-International disseminates microdata encompassing 80% of the world's population, but the coverage for Africa is barely half that, at 42%. Africa is under-represented in the database, not only due to a slow start and ACAP's refusal to cooperate but also because the African statistical offices are exceedingly reluctant to allow outsiders access to the data. Nonetheless, microdata for fifteen African countries (29 censuses, 55 million person records) are currently available and Africa has become a top priority as more African census data are entrusted. Integration work is underway for another fifteen countries, but some very important nations—Nigeria, Algeria, Zimbabwe, etc.—are not yet participating (see Appendix, table A1). Microdata are inaccessible for one-third of the population of Africa.[1]

The African Development Bank, seeking to promote open access to census microdata recently commissioned a technical expert to visit statistical offices that have been slow to open their doors to assemble the data on the spot. Success was achieved within months in five countries—Benin, Cameroon, Ethiopia, Liberia, and Mozambique. Funding remains on the table, awaiting a signal to proceed, for 17—Algeria, Burundi, Central African Republic, Comoros, Congo-Republic, Cote d'Ivoire, Equatorial Guinea, Eritrea, Gabon, The Gambia, Libya, Mauritania, Namibia, Nigeria (National Population Commission), Swaziland, Tunisia, and Zimbabwe. Hopefully the Bank's support will bear fruit in the not too distant future so that the vast major of African statistical offices make their census microdata available.

This paper analyzes 24 African census samples (13 countries) available from the IPUMS website to illustrate how microdata may be used to assess development and pinpoint basic human needs at local administrative levels over time. We calculate a Human Development Index-like measure for small areas (typically municipalities, henceforth denoted as MHDI), recently proposed by Permanyer (2013). Unlike the United Nations Development Program's classic HDI, Permanyer's measure is computed solely from census microdata and therefore, when the data are accessible, may be easily calculated for small administrative areas, where much of the responsibility lies for executing policies related to health, education and general well-being. Summarizing the UNDP's HDI at the national level has its attractions, but the MHDI exposes inequalities exist within country at the same time that it offers a summary statistic for the entire country, although somewhat different from the classic HDI. In this respect, the MHDI is one of the latest attempts to construct human development indicators defined below the country level[2].

One of the most attractive features of the use of complete census data is the possibility of disaggregating national-level averages and exploring the distribution of human development and its components with unprecedented geographical detail. In particular, the availability of complete census microdata allows pinpointing those administrative units leaping ahead or lagging behind in the pace of well-being progress. Therefore, the MHDI methodology can be particularly useful for policy-makers in need of highly detailed data.

---

[1]The Integrated Household Survey Network has supported capacity building in National Statistical Offices to facilitate access to microdata. Unfortunately, while the project appears to be successful—in Africa the number of websites hosted by national statistical offices number in the dozens—as a matter of fact few of these are functional. An exhaustive 2009 study reported that only four African statistical offices provided relatively obstacle free access to microdata (Woolfrey 2009).

[2]Other conceptually related approaches are those of Grimm et al (2008, 2010) who present an HDI for different income quintiles; Harttgen and Klasen (2011a), who calculate the HDI separately for internal migrants and for nonmigrants and Harttgen and Klasen (2011b), who define a household-based Human Development Index. Since these indicators are constructed on the basis of household surveys alone, it is not possible to estimate their distribution in such a way that they are statistically representative for sub-national geographical units (e.g.: state, province, municipality and so on) because of sampling variability.

The MHDI is a composite with three components: health (proportion surviving of live-born children), education (a composite of literacy and primary education completion), and standard of living (assets, such as potable water, waste disposal and electricity). For countries with two or more suitable sets of census microdata, we compare change over time. For all countries with at least one set we offer cross-national comparisons and calibrate the national census-based measure against the conventional HDI.

The paper is structured as follows. In section 2 we present the definitions, the data and the methodology that has been used to construct the MHDI for 24 census samples in 13 African countries. The empirical results of our analysis are shown in section 3. We discuss the implications of our results in section 4. We conclude with a discussion of methodological, theoretical, and policy implications as well as an appeal to African statistical agencies that have not yet done so to open access to census microdata. Despite the pessimism in the epigraph, we argue that Africa is the continent to benefit the *most* from the MHDI, when African census agencies adopt twenty-first century principles of access to microdata.

2. Methodology

2.1. Data

Our analysis is based on harmonized census microdata samples from the Integrated Public Use Microdata Series (IPUMS) international database (Minnesota Population Center 2010). The dataset used here contains 24 samples from 13 countries. These samples were taken from censuses between 1982 and 2008 (see Table 1 for details on the countries and years included in the dataset). Unfortunately, countries available in the IPUMS database like Egypt could not be included in our analysis because they lacked the appropriate child survivorship variables to compute the health component of the MHDI.

The geographical detail available for each country is not uniform, as it depends on the density of the sample size (typically between 5% and 10%), the distribution of the population and the way in which administrative units are defined for each country (see Table 1). For the case of Rwanda data is only available at the first administrative level (i.e.: the Province level), while for Mali and South Africa indicators can be computed at the third administrative level (i.e.: the so-called municipal level). For the remaining samples, indicators can be collected at the second administrative level (the specific name depends on each country). In case the corresponding statistical agencies allowed access to complete census microdata files, it would be possible to reproduce the analysis presented in this paper with much greater geographical detail.

[[[Table 1 around here]]]

2.2. Some basic definitions

In this section we describe the methodology used to define the MHDI. Following Permanyer (2013), the MHDI for administrative unit '*i*' is an average of the health, education and wealth components (denoted as H_i, E_i and W_i respectively), the construction of which is described below.

Health

The health indicator for administrative unit '*i*' will be the percentage of surviving children born to women in that administrative unit between ages 20-39, which will be denoted by P_i. The Health Index H_i is defined as $H_i=(P_i - P_{min})/(P_{max} - P_{min})$, where P_{min}, P_{max} are the minimal and maximal benchmark values. This is the standard normalization methodology used in the construction of the classical HDI. In our empirical results, we have chosen P_{min}=50 and P_{max}=100. The choice of P_{max}=100 is quite uncontroversial, as it is the natural upper bound that would be observed in the absence of child mortality. The choice of P_{min}=50 is slightly arbitrary but it is grounded on the following reasons: i) It is a simple rounded number that involves no truncation of the distribution of the different P_is; (ii) Lower rounded bounds like 25 or 0 would be theoretically feasible but are too far away from the actual values observed in the distribution of the P_is. Analogous criteria have been used in the construction of the HDI when normalizing life expectancy values in the health component of the index. The health index H_i is particularly suitable to estimate health conditions for small size populations. Among others, it has been used to describe socio-demographic conditions of scattered indigenous populations in Latin America (ECLAC 2010).

Education

In the original HDI definition, the education component is defined as (2/3)*ALR+(1/3)GER, where ALR is the Adult Literacy Rate (defined as the percentage of individuals aged 15 or more who are available to read and write) and GER is the Gross Enrolment Ratio (defined as the number of students enrolled in primary, secondary and tertiary levels of education, regardless of age, expressed as a percentage of the population of theoretical school age for the three levels). While the former indicator focuses on all adults, the latter focuses on the population in school ages. Unfortunately, it is not always possible to compute the values of ALR and GER for all countries because of data limitations. In this respect, we have made the following decisions:

i) Whenever ALR was not available in a given country, we have used an alternative definition using the variable 'Years of Schooling'. More specifically, we considered those individuals with less than five years of schooling as being illiterates. This approach has been used among others by Grimm et al (2008, 2010). In order to validate the reasonableness of this approach, we have performed a couple of consistency checks.

Consistency check #1 (macro level check): For each administrative unit where both indicators were available, we have compared their values. It turns out that the correlation coefficient between them within each census sample is extremely high, almost always above 0.95. To illustrate, we show the results at the country level for the 18 samples where both indicators were available at the same time. The results – shown in Figure 1 – indicate that both indicators are highly consistent, with a correlation coefficient equal to 0.96.

[[[Figure 1 around here]]]

Consistency check #2 (micro level check): For each administrative unit where both indicators were available, we have cross-checked the classification of individuals according to both criteria (i.e.: literate/illiterate vs. less than five years of schooling/five years of schooling or more). In most cases, the percentage of agreement between both criteria is above 90%, with a few countries having agreement rates around 85%.

ii) In order to compute GER, it is necessary to have the variable 'School Attendance'. Unfortunately, this indicator is unavailable for 5 of the 24 samples considered in this paper. Unlike the previous case in which it was possible to present an alternative way of defining literacy (see (i)), there is no clear cut way of presenting alternative definitions of GER with the available data. For this reason, and in order to maximize the geographical coverage of our analysis, we have opted for an alternative solution defining a new indicator that is somewhat similar in spirit to GER. If we define $PR+_{15\text{-}24}$ as the population aged 15-24 having at least completed primary education and $POP_{15\text{-}24}$ as the population aged 15-24, we can define the index

$$PR = \frac{PR+_{15-24}}{POP_{15-24}} \qquad [1]$$

While GER compares the number of enrolled students with respect to the population of theoretical school age, PR is simply interpreted as the proportion of population aged between 15 and 24 that has at least completed primary education in the administrative area we are dealing with. Similarly to GER, the new indicator focuses on the young population (as opposed to ALR). There are several reasons why we have opted for such an indicator. First, it is very straightforward and simple to understand. Second, it can be computed in all our samples except for one[3]. Third, it is an indicator that is directly related to the achievement of Millennium Development Goal #2 (ensure universal primary schooling). In those countries where both GER and PR are available, we have compared their values for the different administrative units we are working with. It turns out that the correlation coefficient between them is very high, in most cases around 0.9.[4] Therefore, the administrative unit rankings that are derived from the values of both indicators are very similar. To illustrate, Figure 2 shows the country level values of GER and PR for those census samples where both indicators are available. Again, both indicators are relatively similar and the correlation coefficient is very high (0.88).

---

[3]In the case of Rwanda 1991, that information is not available. In order to include that country in our sample, we have defined a simpler version of the education index, in which we only included the ALR. Since this compromises the comparability of that specific country, special caution should be exercised when interpreting the corresponding results.

[4]The big exception to that rule was found in South Africa, which has the largest levels of school attendance, primary education completion and human development among the countries included in this paper. In that case the correlation coefficient is around 0.3.

[[[Figure 2 around here]]]

Summing up, the education index for administrative unit '*i*' used in this paper can be written as E_i=(2/3)*ALR_i+(1/3)*PR_i.

Standard of living

The standard of living index for administrative unit '*i*' (W_i) is an average of a household asset index defined for all households belonging to '*i*'. Our asset indices are constructed at the household level (*h*) using the following aggregation formula:

$$A_h = \frac{a_{h1} + \ldots + a_{hk}}{k} \qquad [2]$$

where A_h is the asset index for household *h*, the $a_{hj} \in \{0,1\}$ refer to the absence/presence of asset *j* in household *h* and *k* is the number of assets we are taking into account. After computing the asset index A_h for each household in the census, our wealth index (W_i) is computed for each municipality '*i*' as a weighted arithmetic mean of the asset indices of the households belonging to '*i*' (each household weighted by its population share within the corresponding municipality). The availability of household assets questions varies widely across African census samples, an issue that has imposed serious challenges to our effort of developing comparable measures of standard of living across time and space. Given the aforementioned questionnaire variability, we have opted for a two-pronged strategy to maximize the use of data. On the one hand, and in order to ensure international comparability, we have defined a standard of living indicator that included all assets that were available in the different questionnaires *at the same time*. This has produced an extremely crude asset index consisting of three components only: access to clean drinking water, access to electricity and ownership of an improved sanitation facility.[5] Such simple asset index will be referred to as 'core standard of living index' or 'core wealth index' and will be denoted as W_i^{Core}. On the other hand, and given the crudeness of the core standard of living index, we have introduced a country specific asset index – denoted as W_i^{CS} – that might better proxy the wealth distribution in that country. That index has been used for within-country comparisons only. The items that have been introduced in each country specific standard of living index are shown in Table 2. In the construction of the different asset indices we have chosen an equal weighting scheme (as done by many others, e.g.: Montgomery et al 2000, Case, Paxson and Ableidinger 2004, Hohmann and Garenne 2010, Permanyer 2013). This way, the meaning of the indices is crystal clear: they simply count the proportion of owned assets.

[5]As can be seen in Table 2XXX, there are a couple of exceptions to that rule. For the case of Mali there is no information regarding the water supply and for the case of Malawi 1987 there is no information on access to electricity. In those cases, the corresponding asset index is based on the remaining two components only. Since this compromises the comparability of those specific countries, special caution should be exercised when interpreting the corresponding results.

For the sake of completeness, Figure 3 plots the joint values of W_i^{Core} and W_i^{CS} computed at the country level. Interestingly, it seems that both indices tend to line up the African countries included in this study in a quite similar way (the correlation coefficient equals 0.95). Therefore, despite the crudeness of its definition, the values of the core wealth index might not be overly misleading when estimating the underlying wealth distribution.

[[[Figure 3 around here]]]

[[[Table 2 around here]]]

The municipal-based HDI

After computing the three components of the index, the MHDI for administrative unit '*i*' is finally defined as the arithmetic mean (H_i + E_i + W_i)/3. It should be highlighted that since 2010, the official HDI is calculated using the geometric mean $\sqrt[3]{H_i \cdot E_i \cdot W_i}$. Both approaches have their advantages and disadvantages. On the one hand, the multiplicative HDI was introduced to reward countries with balanced (i.e.: similar) distributions across components and penalize those countries with unequal achievements. However, the multiplicative index drops to zero whenever any of its components is equal to zero – regardless of the value of the other two. This problem is more likely to be found when the units of analysis are very small, as it becomes increasingly possible that some components of the index equals zero. On the other hand, the additive HDI is insensitive to the extent to which achievements across dimensions are balanced or not. However, it does not have the boundary problems of its multiplicative version and – importantly for the purposes of this paper – it allows knowing the contribution of the different components to overall inequality in human development, as is shown below.

Inequality decomposition by factor components

Following Permanyer (2013), we briefly present the methodology used in this paper to compute the contribution of the different components to overall inequality in human development. For each administrative unit '*i*' let Y_i, H_i, E_i and W_i be the corresponding human development, health, education and wealth indices. In case of additive human development indices we have that

$$Y_i = \frac{H_i}{3} + \frac{E_i}{3} + \frac{W_i}{3} \qquad [3]$$

The distribution of human development, health, education and wealth indices will be denoted as *Y*, *H*, *E* and *W* respectively. According to Shorrocks (1982:195), if the human development distribution is ordered so that $Y_1 \leq Y_2 \leq \ldots \leq Y_n$, then the corresponding Gini inequality index can be written as

$$G(Y) = \frac{2}{n^2 \mu_y} \sum_{i=1}^{n} \left(i - \frac{n+1}{2} \right) Y_i \qquad [4]$$

where n is the number of administrative units we are taking into account and μ_y is the mean of the human development distribution. Plugging equation [3] into equation [4] we have

$$G(Y) = \frac{2}{n^2 3\mu_y} \sum_{i=1}^{n} \left(i - \frac{n+1}{2} \right) (H_i + E_i + W_i) = \frac{\mu_h}{\mu_y} \overline{G}(H) + \frac{\mu_e}{\mu_y} \overline{G}(E) + \frac{\mu_w}{\mu_y} \overline{G}(W) \qquad [5]$$

where μ_h, μ_e and μ_w are the means of the health, education and wealth distributions and

$$\left. \begin{aligned} \overline{G}(H) &= \frac{2}{3n^2 \mu_h} \sum_{i=1}^{n} \left(i - \frac{n+1}{2} \right) H_i \\ \overline{G}(E) &= \frac{2}{3n^2 \mu_e} \sum_{i=1}^{n} \left(i - \frac{n+1}{2} \right) E_i \\ \overline{G}(W) &= \frac{2}{3n^2 \mu_w} \sum_{i=1}^{n} \left(i - \frac{n+1}{2} \right) W_i \end{aligned} \right\} \qquad [6]$$

which are known as the pseudo-Ginis for factors *H*, *E* and *W* respectively (see Shorrocks 1982:196 and Lerman and Yitzhaki 1985:152). Equation [5] shows a natural additive decomposition of the Gini index where the contribution of the *H*, *E* and *W* components is clearly established.

3. Empirical results

In this section we present the empirical findings of the paper regarding the MHDI distribution across 13 African countries. We start exploring distributions within countries first and then proceed with comparisons between countries. In addition, we will put into practice the inequality decomposition by factor components methodology presented at the end of section 2.

3.1. Within country analysis.

When the health, education, wealth and human development indicators (i.e.: the H_i, E_i, W_i and $MHDI_i$) are available for each administrative unit we are working with, it becomes possible to explore their distribution across the entire country with great geographical detail. In addition, when more than one census is available for the same country, it is particularly interesting to investigate the evolution of the human development distribution and its three components over time. In our study, there are eight countries with more than one census (Kenya, Malawi, Mali, Morocco, Rwanda, South Africa, Tanzania and Uganda). The results are shown in Figure 4: for each of those countries we plot the distributions of human development together with the health, education and wealth components across the corresponding administrative units. Since comparisons in this subsection are within countries we will use the country-specific standard of living indices (W_i^{CS}), which are expected to better capture the underlying wealth distribution than the one derived from the values of the core indicator W_i^{Core}.[6] Therefore, it is important to highlight that while the distribution of the health and education components shown in Figure 4 are comparable across countries, the wealth and overall human development distributions are not. The corresponding comparable results across countries will be presented in section 3.2.

[[[Figure 4 around here]]]

As can be seen from Figure 4, most distributions have the expected direction of improvement over time, an encouraging result for the corresponding countries. However, there are important exceptions to these encouraging trends. The Rwandan health distribution deteriorates from 1991 to 2002, a phenomenon that can be attributed to the massive killings that took place in the country in 1994. As a consequence, the overall MHDI distribution does not show signs of clear improvement either (despite improvements in the EI and WI distributions). In addition, the health distribution reported for South Africa also deteriorates from 2001 to 2007, a result that is in line with the decreasing official figures of life expectancy reported in that country and which can be attributed to a large extent to the high prevalence of HIV/AIDS.

Inspecting the shape of the density functions shown in Figure 4, one can see that there are important variations across countries. Rather than observing the traditional unimodal and highly skewed shapes that characterize income distributions, many of the distributions shown in Figure 4 are very bumpy. This suggests that the levels of inequality and polarization in those countries can be very large, an issue that will be explored in more detail below. We hypothesize that the existence of bumps in those distributions might be attributable to the urban - rural divide: urban households tend to own more assets and their inhabitants are more likely to enjoy the benefits of nearby health and education facilities. Finally, if one compares the spread of the distributions within countries over time, no substantial changes seem discernible at first sight. At the end of the following section, we will quantify precisely the extent of inequality in all those distributions.

3.2. Between country analysis.

We will now compare the distribution of human development and its components across countries. For that purpose, we will make use of the 'core' wealth index W_i^{Core} that

[6]It should be pointed out that for the cases of Malawi and South Africa, we only show the results corresponding to their last two censuses. Both countries have a third census which, unfortunately, does not have the same list of household assets as the other two, so they are not strictly comparable. Morocco is the only country with the same list of variables for the three available censuses, so the corresponding results are shown for all of them.

includes the same assets for all the countries included in this study – thus ensuring cross-country comparability. We start examining the population weighted country-level average of our MHDI indicator and its health, education and standard of living components, which are shown in Table 3. The country average MHDI values differ greatly, ranging from 0.174 (observed in Mali 1987) to 0.837 (South Africa 2007). Being an average of three different components, it is also important to explore them separately. Interestingly, the three components of the MHDI behave quite differently. The values of the country average health index range from 0.308 (Mali 1987) to 0.899 (South Africa 2001), those of the education index from 0.168 (Mali 1998) to 0.875 (South Africa 2007) and those of the core wealth index from an appallingly low 0.016 (Uganda 1991) to 0.752 (South Africa 2007). Taking into account the fact that the theoretical range of these indicators is the interval [0,1], the observed range of variation for each case is considerably large. This illustrates the heterogeneity that is observed between the African countries included in the analysis.

Since the MHDI is simply the arithmetic mean of the HI, EI and WI indices, it is straightforward to compute the contribution of each of these subcomponents to the aggregate value of the index. To illustrate: the percentual contribution of the health component to the aggregate MHDI value is simply computed as 100*HI/(HI+EI+WI). As is shown in Figure 5, the contribution of the three components varies greatly across countries. Figure 5 shows that as the country-level MHDI values decrease, the relative contribution of the wealth index tend to decrease as well while the contribution of the health component tends to increase. As can be seen, the percentage contribution of the three components is balanced (i.e.: around 33% each) only for those countries with the largest MHDI values (South Africa and Morocco). For the other countries, the MHDI values tend to be overwhelmingly accounted for by the health and education components.

[[[Table 3 around here]]]

[[[Figure 5 around here]]]

In order to contrast the results of our methodology with the official HDI results reported yearly in the Human Development Reports, the latter are also shown in Table 3. Figure 6 plots the country-level values of our MHDI indicator against UNDP’s HDI. The results of this validation exercise are quite encouraging: the values of the MHDI and HDI are closely related in a linear fashion, and no large discrepancies are observed (the correlation coefficient for the values plotted in Figure 6 is very high: 0.94). The country that differs the most from the predicted linear model is Rwanda 1991, perhaps because it is the only country where the education component has been calculated using a slight variation with respect to the others that might have artificially inflated its ‘true’ value. As can be seen, the values of MHDI tend to be higher than those of the official HDI (the dots are mostly below the dashed equality line). This issue is not particularly troubling since neither the HDI nor the MHDI have a specific unit of measurement. Therefore, what is especially relevant is the ordinal information (i.e.: the rankings) that is derived from the values of those indices, rather than the cardinal values themselves.

[[[Figure 6 around here]]]

The country-level MHDI values shown in Table 3 and Figures 5, 6 are the result of averaging MHDI values across a large amount of administrative units defined at sub-national level. Despite the interest that such country-level averages might have, the main rationale for introducing the MHDI methodology is to uncover the inequalities that are hidden behind those aggregate numbers. In order to compare not only the average value of the MHDI distribution but also its spread within the corresponding country, Figure 7 plots the density functions of the MHDI distributions for all African countries included in the analysis except for the case of Sudan (its 2008 census was conducted about a decade later than the others, an issue that seriously compromises its comparability). For those countries with several observations, we have chosen those belonging to the 2000 census round. As can be seen, there are huge variations not only in the average MHDI values but also in the spread of the human development distributions within countries[7]. Again, the shape of some of these distributions is very bumpy, therefore suggesting that the levels of inequality and polarization in the corresponding countries must be very high (this is particularly the case for Guinea, Malawi, Mali, Rwanda and Sierra Leone). At the other extreme, countries like Kenya, Morocco and South Africa have relatively smooth distributions with a relatively wide range of variation. In this context, it is particularly interesting for policy making purposes to identify the administrative units that are located in the lower and upper tails of the corresponding MHDI distributions. In the following section, we quantify more precisely the extent of inequality observed in these distributions.

[[[Figure 7 around here]]]

Inequality in human development

Table 4 shows the values of the Gini index for the MHDI, HI, EI and WI distributions. For the case of the MHDI distribution, its values range from 0.055 (observed in Rwanda 1991) to 0.274 (Mali 1987). This range of variation is relatively similar to the one observed for the Gini index of the health distributions: from 0.032 (observed in South Africa 2001) to 0.206 (Mali 1987). However, the range of variation is larger for the Gini index of the education distributions (from 0.042 in South Africa 2007 to 0.401 in Mali 1998) and even larger for the wealth distributions (from 0.155 in South Africa 2007 to 0.825 in Mali 1987). Taking into account the fact that the Gini index can only take values between 0 and 1 (0 denoting complete equality and 1 complete inequality), the observed Gini values for many of the wealth distributions are appallingly high[8]. This

[7] It is important to highlight that the level of geographical disaggregation is not the same for all countries (see Table 1XXX). Therefore, countries with greater geographical detail are likely to exhibit larger spread in their MHDI distribution. This should be borne in mind when comparing distributions' spread.

[8]As mentioned in the previous footnote, when interpreting the values of those Gini indices it should be borne in mind that the level of geographical disaggregation is not the same across countries. It is to be

suggests that the three assets included in the core wealth index (access to clean drinking water, access to electricity and ownership of an improved sanitation facility) are very unevenly distributed within these countries (most probably concentrated in metropolitan urban areas). All in all, the country with highest human development inequality levels in our sample is Mali (in 1987). At the other extreme, South Africa is among the countries with lowest observed levels of inequality in human development.

Table 4 also shows the results of the inequality decomposition by factor components presented at the end of section 2. As can be seen, the contributions of the three components to MHDI inequality do not show clear cut patterns. An inspection of the values of the component-specific Gini indices on the one hand and the contribution of that component to the overall inequality in human development on the other reveals that both magnitudes do not necessarily run in the same direction. In other words: high values of a component-specific Gini index do not necessarily imply that the corresponding contribution to overall inequality in human development is also high (e.g.: Mali has a very high Gini index for the wealth component but the percentual contribution of that component to overall inequality in human development barely reaches 20%). This apparently surprising fact is attributable to the weak correlation structure of the data (i.e.: the administrative units' rankings within a given country can be quite different depending on the specific indicators that are used to rank them, see Shorrocks 1982 and Lerman and Yitzhaki 1985 for details).

As is shown in Table 4, the health component is the one that tends to contribute the least to observed MHDI inequality levels, but there are important exceptions (Mali 1987 and 1998, Rwanda 1991 and 2002). The education and wealth components tend to dominate the contribution to overall MHDI inequality, but, again, it is difficult to discern simple patterns in the data. Comparing the results of Table 4 with those of Table 3, it turns out that the contribution of the wealth component to overall human development inequality tends to be larger for countries with higher human development *levels*. For instance, South Africa and Morocco, the countries with highest human development levels in our sample, are the countries where the contribution of the wealth component to overall inequality is the largest (above 50%). At the other extreme, in Mali, Rwanda and Uganda (the countries with lowest human development levels in our sample), the contribution of the wealth component is at its lowest values (around 20%). These results, together with the examination of the component-specific distribution graphs shown in Figure 4, lead us to hypothesize that as countries progress towards higher human development levels, the education and health distributions tend to become more homogeneous, therefore increasing the contribution of the wealth component to overall

---

expected that the reported Gini indices can be underestimated for those countries with little geographical detail (e.g.: Rwanda or Sierra Leone).

human development inequality[9]. However, this is a difficult issue beyond the scope of this paper that should be carefully explored in future research.

[[[Table 4 around here]]]

4. Conclusions

In this paper we propose using new measurement techniques recently proposed by Permanyer (2013) which might greatly contribute to add human development indices at very small aggregation levels as an operational toolkit that can be widely used by scholars, researchers, practitioners, national and international institutions and policy makers alike. As argued throughout the paper, highly disaggregated information is extremely important for a variety of purposes ranging from academic research to the design of development policies. On the academic side, the lack of reliable data at sub-national levels is a major hurdle that critically undermines the possibility of (i) assessing the large unmeasured heterogeneity within countries; and (ii) empirically testing alternative theoretical efforts proposed in different disciplines of the social sciences that aim to establish formal linkages and interactions between variables operating at the micro and macro aggregation levels. From the policy-making perspective, there is a huge need for more accurate information that can be used for the design and evaluation of public policy and to reduce the risk of falling into the ecological fallacy. The design of fine-tuned policy instruments can be particularly useful to locate and monitor the evolution of small administrative units that are otherwise concealed under national averages.

The methodological proposals introduced in this paper are particularly pertinent for the case of Africa. While Devarajan (2013) and Jerven (2013) are correct when they conclude that data for much of Africa is not appropriate for economic growth rates or per capita income estimates, the MHDI analysis shown in this paper illustrates that they are good enough for many other purposes. Indeed, a major aggravating problem that contributes to the 'African statistical tragedy' is the lack of accessibility to existing data. This paper aims to illustrate the usefulness of census data – which are vastly underutilized in Africa (Alderman et al 2003, p.193) – and hopefully contribute to make them transparent and freely accessible.

References

Alderman, H., Babita, M., Demombynes, G., Makhatha, N., & Özler, B. (2003). How Low Can You Go? Combining Census and Survey Data for Mapping Poverty in South Africa. *Journal of African Economies*, *11*(2), 169-200.

Case, A., Paxson C., & Ableidinger, J. (2004). Orphans in Africa: Parental Death, Poverty, and School Enrollment. *Demography, 41*(3), 483-508.

[9] Albeit in a completely different geographical context, the results shown in Permanyer (2013) for Mexico are in line with this hypothesis.

Devarajan, Shantayanan. (2013). Africa's Statistical Tragedy. *Review of Income and Wealth* 2013. DOI: 10.1111/roiw.12013

ECLAC (2010). *Pueblos Indígenas y Afrodescendientes*. Publication series by ECLAC's Population Division.

Grimm, M., Harttgen, K., Klasen, S. & Misselhorn, M. (2008). A Human Development Index by Income Groups. *World Development*, *36* (12), 2527-2546.

Grimm, M., Harttgen K., Klasen, S., Misselhorn, M., Munzi, T. & Smeeding, T. (2010). Inequality in Human Development: An Empirical Assessment of 32 Countries. *Social Indicators Research*, *97*, 191-211.

Harttgen, K. & Klasen, S. (2011a). A Human Development Index by Internal Migration Status. *Journal of Human Development and Capabilities*, *12*(3), 393-424.

Harttgen, K. & Klasen, S. (2011b). A Household-based Human Development Index. *World Development*, *40*(5), 878-899.

Hohmann, S. & Garenne, M. (2010). Health and wealth in Uzbekistan and sub-Saharan Africa in comparative perspective. *Economics and Human Biology*, *8*(3), 346-360.

Jerven, M. (2013). *Poor Numbers: How We Are Misled by African Development Statistics and What To Do About It*. Ithaca: Cornell University Press.

Lerman, R. & Yitzhaki, S. (1985). Income inequality effects by income sources: a new approach and applications to the United States. *The Review of Economics and Statistics*, *67* (1), 151-156.

Minnesota Population Center. 2010. Integrated Public Use Microdata Series, International; Version 6.0 [Machine-readable database]. Minneapolis, University of Minnesota.

Montgomery, M., Gragnolati, M., Burke, K., & Paredes, E. (2000). Measuring Living Standards with Proxy Variables. *Demography 37*(2), 155-174.

Permanyer, I. (2013). Using Census Data to Explore the Spatial Distribution of Human Development. *World Development* (DOI:10.1016/j.worlddev.2012.11.015).

Shorrocks, A. (1982). Inequality decomposition by factor components. *Econometrica*, *50*(1), 193-211.

Woolfrey, L. (2009). African microdata access survey. International Statistical Institute World Congress, Invited Paper Session 100, Durban, South Africa.

Appendix
Table A1

IPUMS-Africa, February 2013

Inventory by Country of Censuses, Microdata Entrusted and Needed

	IPUMS-International		**Census Round**					
Country	**Entrusted**	**Need**	**2005-14**	**1995-04**	**1985-94**	**1975-84**	**1965-74**	**1955-64**
A. Microdata for 15 countries, 41 sets of census microdata entrusted, 29 integrated and disseminating								
Egypt	**3**	Integrate 1986	**2006**	**1996**	**1986**	1976	1966	
Ghana	**2**	**Need 2010**	**2010**	**2000**		**1984**	1970	
Guinea	**2**	-		**1996**		**1983**		
Kenya	**5**	Integrating 2009	**2009**	**1999**	**1989**	**1979**	**1969**	
Malawi	**3**	Completed	**2008**	**1998**	**1987**	1977	1966	
Mali	**3**	Integrating 2009	**2009**	**1998**	**1987**	**1976**		
Morocco	**3**	Completed	2014	**2004**	**1994**	**1982**	1971	1960
Rwanda	**2**	Completed	2012	**2002**	**1991**	1978		
Senegal	**3**	Integrate 1976	2012	**2002**	**1988**	**1976**		
Sierra-Leone	**1**	Completed	2014	**2004**	1985		1974	1963
South-Africa	**6**	**Need 2011**	2011/**7**	**2001**	**1996**	**1980, 5**	**1970**	1960
South-Sudan	**1**	Completed	**2008**					
Sudan	**4**	Integrate 73-93	**2008**		**1993**	**1983**	**1973**	
Tanzania	**2**	**Need 2012**	2012	**2002**		**1988**	1978	1967
Uganda	**2**	**Need 2013**	2013	**2002**	**1991**	1980		1969
B. 13 Countries (29 datasets) entrusted to IPUMS-International. Integrating (bold); awaiting most recent census								
Benin	**3**	**Need 2013**	2013	**2002**	**1992**		**1979**	1961
Botswana	**3**	**Need 2011**	**2011**	**2001**	**1991**	**1981**	1971	1964
Burkina-Faso	**3**	Launch: 2013	**2006**	**1996**	**1985**	1975		
Cameroon	**3**	Launch: 2013	**2005**			**1987**	**1976**	
Chad	**1**	**Need 2009**	**2009**		**1993**			1962
Ethiopia	**3**	Launch: 2014	**2007**		**1994**	**1984**		
Lesotho	**2**	Launch: 2014	**2006**	**1996**	1986	1976	1966	
Liberia	**2**	Launch: 2014	**2008**				**1974**	
Madagascar	**1**	**Need 2013**	2013		**1993**		1975	1966
Mauritius	**2**	**Need 2010**	2010	**2000**	**1990**	**1983**	1972	1962
Mozambique	**2**	Launch: 2014	**2007**	**1997**		1980	1970	1960
Niger	**2**	**Need 2009**	**2012**	**2001**	**1988**	1977		
Zambia	**2**	**Need 2010**	**2010**	**2000**	**1990**	1980		1969**?**
C. Agreement signed, but no microdata sets entrusted: 4 countries								
Cape Verde	**0**	**Need all**	**2010**	**2000**	**1990**	1980		
Central-African- Rep.	**0**	**Need all**	2013	**2003**		**1988**	1975	1960
Cote-d'Ivoire	**0**	**Need all**	2013	**1998**	**1988**	1975		
Guinea-Bissau	**0**	**Need all**	**2009**		1991		1979	1960
D. No agreement signed; little microdata entrusted: 16 countries								
Algeria	**0**	**Need all**	**2008**	**1998**	**1987**	1977	1966	
Burundi	**0**	**Need all**	**2008**		1990	1979	1970	
Comoros	**0**	**Need all**	2013	**2003**	**1991**	1980		
Congo-Republic	**0**	**Need all**		1996		1984	1974	1960/61
Equatorial Guinea	**0**	**Need all**	2013	**2002**	1994	1983		
Gabon	**0**	**Need all**	2013	2003	1993			
Gambia-The	**2**	**Need all**	2013	**2003**	**1993**	**1983**	**1973**	1963
Libya	**0**	**Need all**	2005	1995		1984	1973	1964
Mauritania	**0**	**Need all**	2012	**2000**	**1988**	1977		
Namibia	**0**	**Need all**	**2011**	**2001**	**1991**	1981	1970	1960
Nigeria	**0**	**Need all**	**2006**		**1991**	1980	1973	1963
São Tomé and Principe	**0**	**Need all**	**2012**	2001	1991	1981		
Swaziland	**1**	**Need all**	**2007**	1997	**1986**	**1976**	1966	
Togo	**3**	**Need 2010**	2010			**1981**	**1970**	**1958**
Tunisia	**0**	**Need all**	2014	**2004**	**1994**	**1984**	**1975**	1966
Zimbabwe	**0**	**Need all**	2012	**2002**	**1992**	1982		1969
E. No census microdata until next census is conducted: 5 countries								

Angola	**0**	**Endorse MOU**	2014			1984	1970	1960
Congo-Democratic Rep.	**0**	**Endorse MOU**	2014			1984	1970	
Djibouti	**0**	**Endorse MOU**						
Eritrea	**0**	**Endorse MOU**						
Somalia	**0**	**Endorse MOU**			1987	1975		

Tables

Country	YEAR	Population	Sample density	Adm. Unit Name	Adm. Level	N. of Adm. Units
Ghana	2000	18941330	10	Districts	2	110
Guinea	1996	7290710	10	Prefectures	2	34
Kenya	1989	21481960	5	Districts	2	31
Kenya	1999	28150940	5	Districts	2	69
Malawi	1987	7986690	10	Districts	2	24
Malawi	1998	9913930	10	Districts	2	26
Malawi	2008	13419770	10	Districts	2	31
Mali	1987	7853840	10	Comunnes	3	221
Mali	1998	9913300	10	Comunnes	3	221
Morocco	1982	20257460	5	Province/prefecture	2	63
Morocco	1994	25880520	5	Province/prefecture	2	60
Morocco	2004	29654400	5	Province/prefecture	2	60
Rwanda	1991	7429180	10	Province	1	11
Rwanda	2002	8433920	10	Province	1	12
Senegal	2002	9945620	10	Department	2	34
Sierra Leone	2004	4942980	10	Districts	2	14
South Africa	1996	40578357	10	Municipalities	3	284
South Africa	2001	44769106	10	Municipalities	3	225
South Africa	2007	47173595	2	Municipalities	3	225
Sudan	2008	38206344	15	Districts/counties	2	202
Uganda	1991	16598197	10	Districts	2	113
Uganda	2002	24974490	10	Districts	2	129
Tanzania	1988	23145678	10	Districts	2	39
Tanzania	2002	33505374	10	Districts	2	56

Table 1. Basic information on the samples included in the analysis. Source: IPUMS database.

		Core vars.			Country-specific variables													
Country	Year	El	WS	Tl	Sw	FC	FH	FL	RO	WL	RF	TV	RD	PC	PH	AU	CL	Tot
Ghana	2000	1	1	1		1		1	1	1								7
Guinea	1996	1	1	1														3
Kenya	1989	1	1	1	1	1		1	1	1								8
Kenya	1999	1	1	1	1	1		1	1	1								8
Malawi	1987		1	1		1							1					4
Malawi	1998	1	1	1		1							1					5
Malawi	2008	1	1	1		1							1					5
Mali	1987	1		1		1		1	1	1								6
Mali	1998	1		1		1		1	1	1								6
Morocco	1982	1	1	1														3
Morocco	1994	1	1	1														3
Morocco	2004	1	1	1														3
Rwanda	1991	1	1	1		1		1	1	1			1					8
Rwanda	2002	1	1	1		1		1	1	1			1					8
Senegal	1988	1	1	1	1			1	1	1	1	1			1			10
Senegal	2002	1	1	1	1			1	1	1	1	1			1			10
S.Leone	2004	1	1	1		1		1	1	1	1	1		1	1	1	1	13
S.Africa	1996	1	1	1		1	1				1	1	1	1	1		1	11
S.Africa	2001	1	1	1		1	1				1	1	1	1	1		1	11
S.Africa	2007	1	1	1		1	1				1	1	1	1	1		1	11
Sudan	2008	1	1	1		1					1	1	1	1	1	1	1	11
Tanzania	1988	1	1	1														3
Tanzania	2002	1	1	1														3
Uganda	1991	1	1	1		1		1	1	1								7
Uganda	2002	1	1	1		1		1	1	1								7

Table 2. Variables used in the construction of the country-specific wealth indices W_i^{CS}. El= 'Electricity', WS= 'Water Supply', Tl='Toilet', Sw= 'Sewage', FC= 'Cooking fuel', FH= 'Heating fuel', FL='Floor', RO= 'Roof', WL= 'Wall', RF= 'Refrigerator', TV= 'Television', RD= 'Radio', PC= 'Personal Computer', PH = 'Phone', AU= 'Autos', CL= 'Cell Phone'. Source: IPUMS data.

Country	Year	HI	EI	WI	MHDI	UNDP's HDI
Ghana	2000	0.599	0.599	0.307	0.502	0.431
Guinea	1996	0.463	0.210	0.105	0.259	NA
Kenya	1989	0.627	0.703	0.170	0.500	0.437
Kenya	1999	0.682	0.662	0.180	0.508	0.424
Malawi	1987	0.375	0.384	*0.136*	0.298	0.274
Malawi	1998	0.449	0.520	0.099	0.356	0.344
Malawi	2008	0.506	0.609	0.099	0.405	0.366
Mali	1987	0.308	0.186	*0.026*	0.174	0.167
Mali	1998	0.416	0.168	*0.058*	0.214	0.245
Morocco	1982	0.665	0.322	0.392	0.460	0.351
Morocco	1994	0.716	0.410	0.540	0.555	0.450
Morocco	2004	0.734	0.524	0.672	0.643	0.530
Rwanda	1991	0.489	*0.541*	0.082	0.371	0.215
Rwanda	2002	0.423	0.498	0.095	0.339	0.301
Senegal	2002	0.684	0.405	0.438	0.509	0.369
S. Leone	2004	0.427	0.350	0.098	0.292	0.286
S. Africa	1996	0.692	0.783	0.629	0.702	0.634
S. Africa	2001	0.899	0.806	0.704	0.803	NA
S. Africa	2007	0.837	0.875	0.752	0.821	0.590
Sudan	2008	0.744	0.439	0.197	0.460	0.373
Tanzania	1988	0.543	0.646	0.152	0.447	0.329
Tanzania	2002	0.603	0.691	0.157	0.484	0.347
Uganda	1991	0.490	0.524	0.016	0.343	0.281
Uganda	2002	0.696	0.632	0.070	0.466	0.365

Table 3. Official HDI, country-level MHDI and its health, education and standard of living components calculated from IPUMS census samples. The values of EI for Rwanda 1991 and the values of WI for Mali and Malawi 1987 are not strictly comparable, as slightly different definitions of these indices have been used for those countries (they are written in italics to distinguish them from the other values). NA = ‘Not available’. Source: Authors’ calculations using IPUMS data.

Country	Year	G_MHDI	G_H	G_E	G_W	%C_H	%C_E	%C_W
Ghana	2000	0.160	0.056	0.181	0.409	9.69	42.64	47.67
Guinea	1996	0.222	0.096	0.329	0.689	23.27	36.12	40.61
Kenya	1989	0.132	0.103	0.130	0.487	29.55	36.48	33.97
Kenya	1999	0.134	0.096	0.136	0.505	26.4	38.55	35.05
Malawi	1987	0.148	0.118	0.156	0.391	30.32	39.2	30.48
Malawi	1998	0.108	0.068	0.122	0.359	21.52	53.17	25.31
Malawi	2008	0.122	0.085	0.108	0.550	25.42	41.6	32.98
Mali	1987	0.274	0.206	0.388	0.825	42.03	42.73	15.24
Mali	1998	0.239	0.160	0.401	0.608	42.18	36.3	21.52
Morocco	1982	0.185	0.074	0.275	0.351	14.64	33.17	52.19
Morocco	1994	0.142	0.046	0.207	0.253	8.94	35	56.06
Morocco	2004	0.108	0.046	0.150	0.164	12.12	37	50.88
Rwanda	1991	0.055	0.047	0.054	0.320	34.44	36.52	29.04
Rwanda	2002	0.071	0.071	0.060	0.338	34.22	37.68	28.1
Senegal	2002	0.198	0.077	0.345	0.371	16.71	29.86	53.43
S.Leone	2004	0.176	0.103	0.220	0.486	24.57	47.53	27.9
S.Africa	1996	0.126	0.087	0.075	0.266	20.6	20.37	59.03
S.Africa	2001	0.085	0.032	0.072	0.185	12.1	27.47	60.43
S.Africa	2007	0.068	0.034	0.042	0.155	13.1	20	66.9
Sudan	2008	0.188	0.070	0.302	0.571	10.03	47.19	42.78
Tanzania	1988	0.093	0.085	0.079	0.400	25.67	33.53	40.8
Tanzania	2002	0.116	0.087	0.109	0.517	20.08	40.2	39.72
Uganda	1991	0.114	0.079	0.119	0.795	27.63	49.17	23.2
Uganda	2002	0.091	0.053	0.112	0.628	17.52	53.54	28.94

Table 4. Gini index for the MHDI, HI, EI and WI distributions. Percent contribution of the health, education and standard of living components to inequality in the MHDI distribution. Source: Authors’ calculations using IPUMS data.

Figures

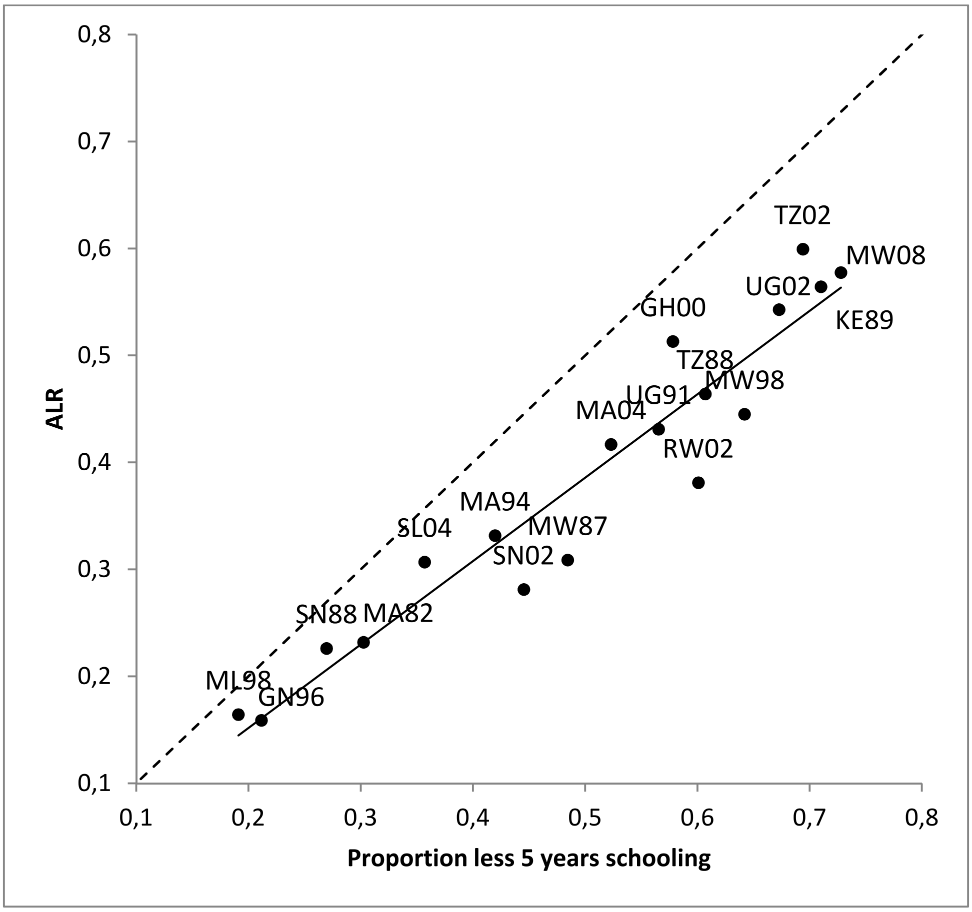

Figure 1. Comparison of ALR (vertical axis) and proportion of adults with less than five years of schooling at the country level (horizontal axis). The dashed line is the 45º equality line. The solid line is the best linear fit line. Countries are labeled with the ISO 3166 codes plus the last two digits of the year in which the corresponding census was conducted. Authors' calculations using IPUMS microdata for 18 African censuses.

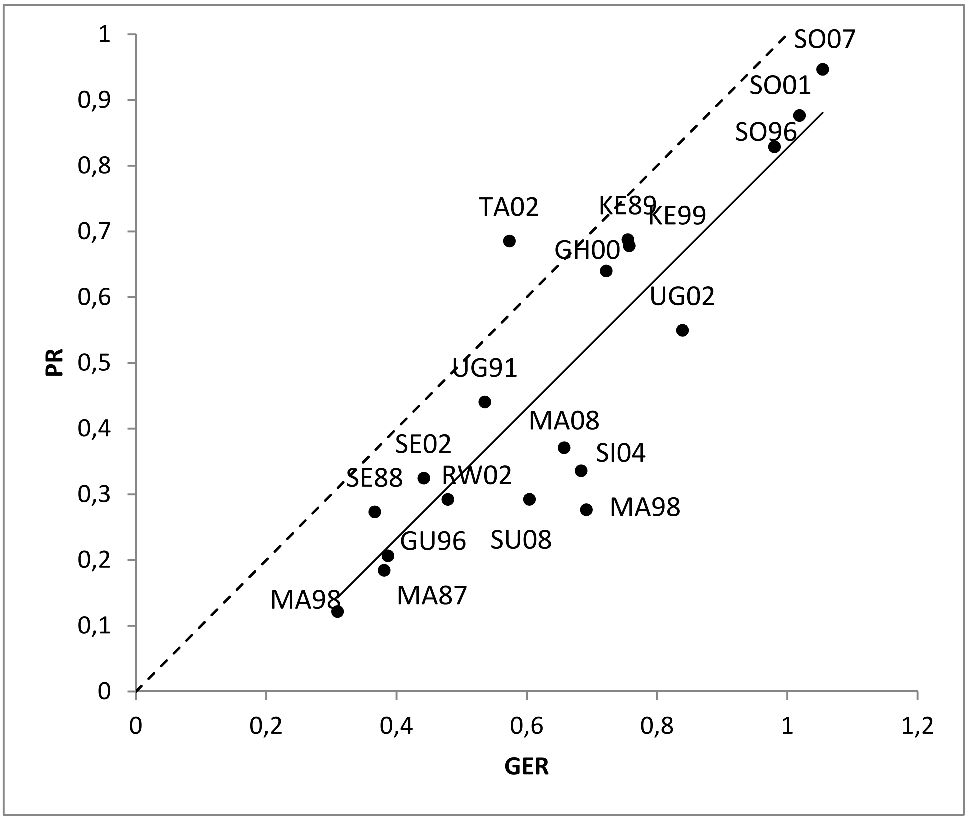

Figure 2. Comparison of Gross Enrolment Ratio (horizontal axis) and Primary completion rates (vertical axis). The dashed line is the 45º equality line. The solid line is the best linear fit line. Countries are labeled with the ISO 3166 codes plus the last two digits of the year in which the corresponding census was conducted. Authors' calculations using IPUMS data.

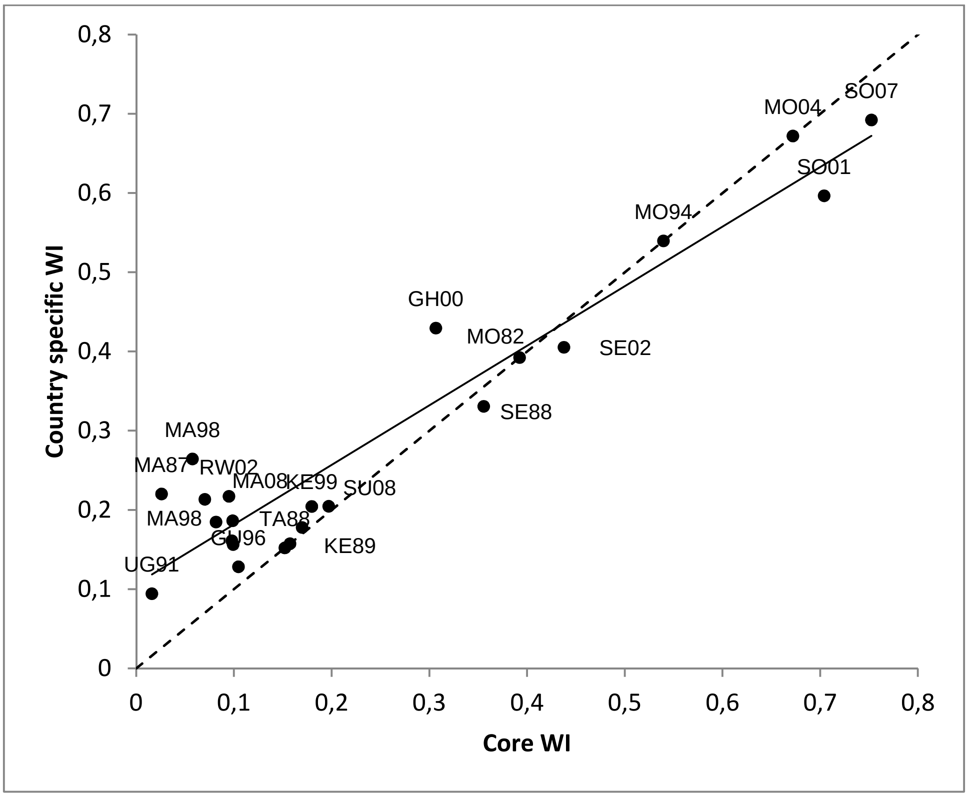

Figure 3. Comparison of the core Wealth Index (horizontal axis) versus the country specific Wealth Index (vertical axis). The dashed line is the 45º equality line. The solid line is the best linear fit line. Countries are labeled with the ISO 3166 codes plus the last two digits of the year in which the corresponding census was conducted. Authors' calculations using IPUMS data.

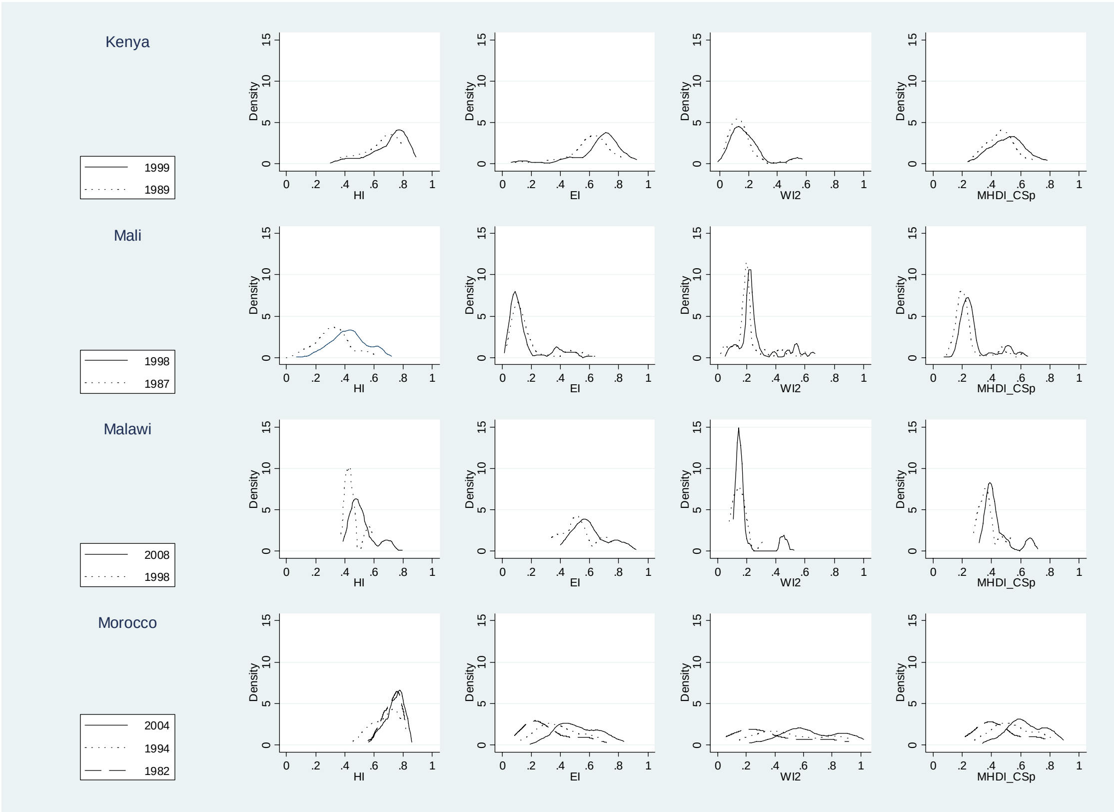

Figure 4. Density functions of the health, education, wealth and human development distributions for Kenya, Mali, Malawi and Morocco. The wealth and human development distributions have been constructed using country specific definitions. Authors' calculations using IPUMS data.

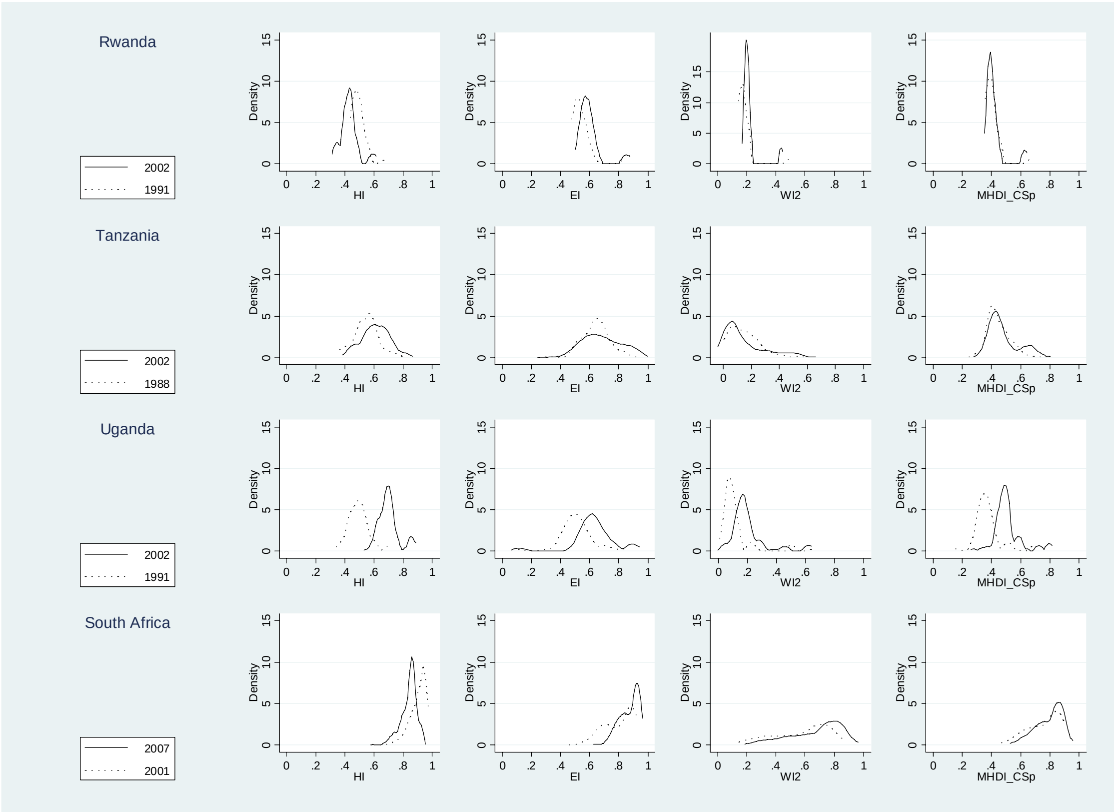

Figure 4 (Continued). Figure 4. Density functions of the health, education, wealth and human development distributions for Rwanda, Tanzania, Uganda and South Africa. The wealth and human development distributions have been constructed using country specific definitions. Authors' calculations using IPUMS data.

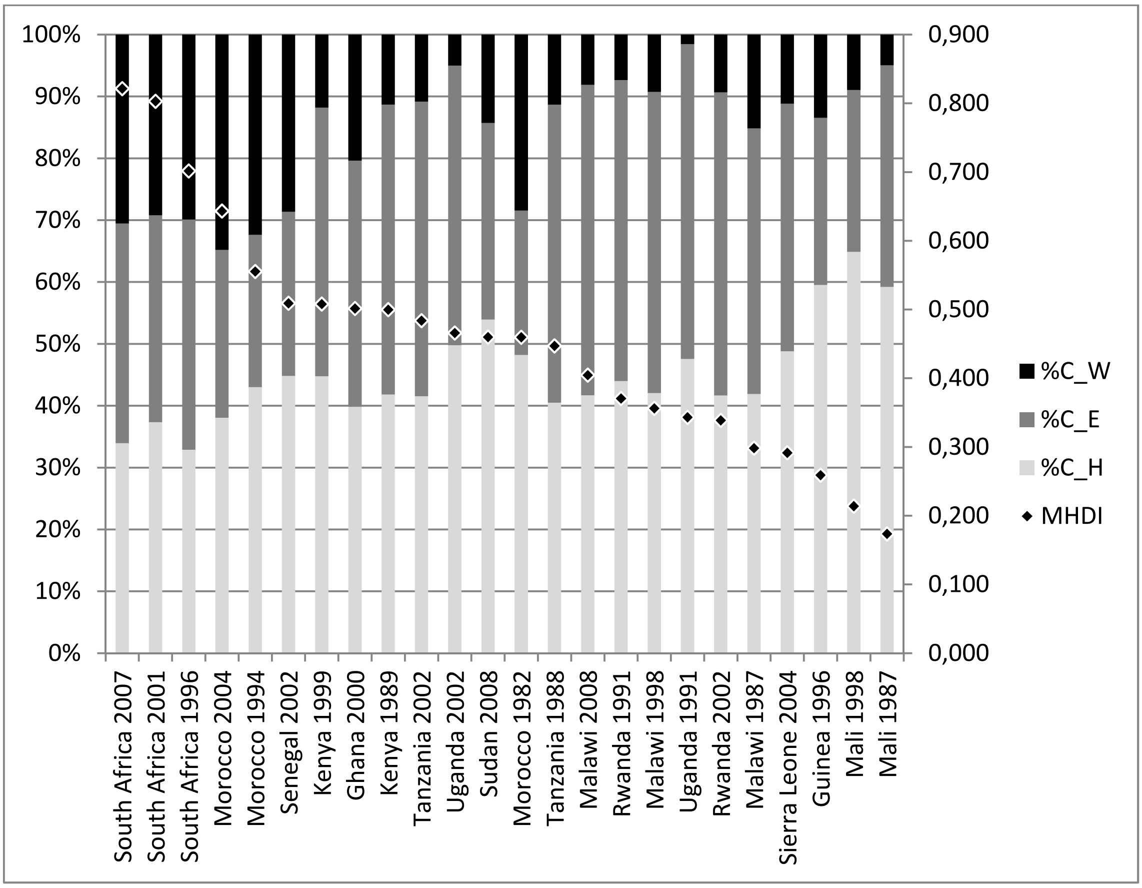

Figure 5. Country-level MHDI values (right vertical axis) with the corresponding percent contributions of the health, education and standard of living components (left vertical axis). Source: Authors' calculations using IPUMS data.

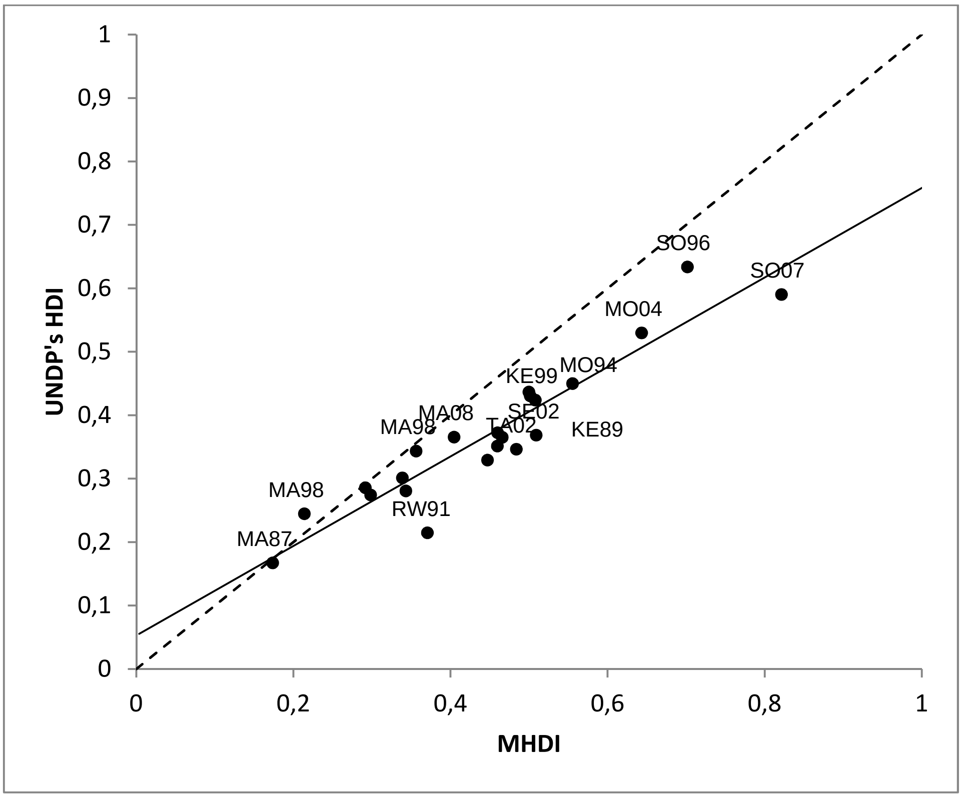

Figure 6. Country level MHDI (horizontal axis) vs UNDP's HDI values (vertical axis). The dashed line is the equality line. The solid one is the best linear fit line. Countries are labeled with the ISO 3166 codes plus the last two digits of the year in which the corresponding census was conducted. Source: Authors' calculations using IPUMS and HDRs data.

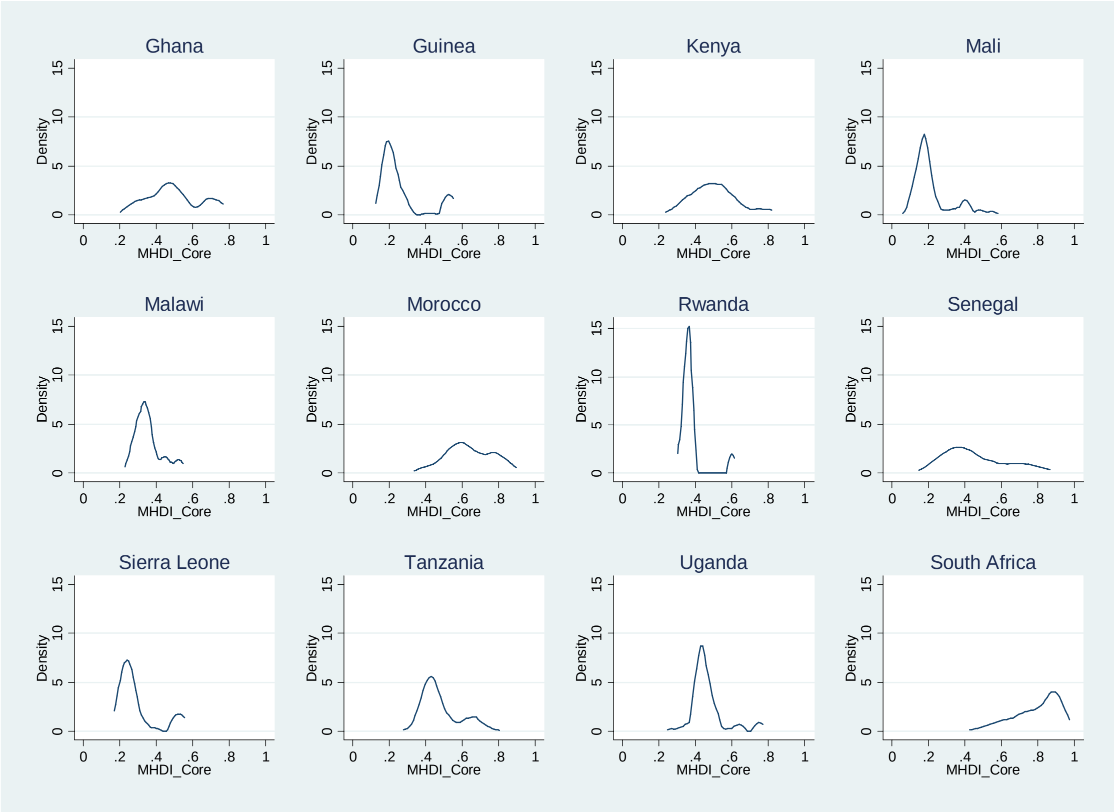

Figure 7. Density functions of the MHDI distributions for the 12 African countries included in the analysis around year 2000 (Ghana 2000, Guinea 1996, Kenya 1999, Malawi 1998, Mali 1998, Morocco 2004, Rwanda 2002, Senegal 2002, Sierra Leone 2004, South Africa 2001, Tanzania 2002, Uganda 2002). Source: Authors' calculations using IPUMS data.