

Harttgen, Kenneth; Klasen, Stephan

**Conference Paper**

## A household-based Human Development Index

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session:  
Empirical Studies of Inequality, No. E4-V2

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Harttgen, Kenneth; Klasen, Stephan (2010) : A household-based Human Development Index, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Empirical Studies of Inequality, No. E4-V2, Verein für Socialpolitik, Frankfurt a. M.

This Version is available at:

<https://hdl.handle.net/10419/37505>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# A household-based Human Development Index

**Kenneth Harttgen and Stephan Klasen\***

*University of Göttingen, Department of Economics*

January 19, 2010

## **Abstract**

The most limiting weakness of the human development index (HDI) is that it looks only at average achievements and does not take into account the distribution of human development within a country or population subgroups. All previous attempts to capture inequality in the HDI have also used aggregate information and there exists no HDI at the household level. This paper provides a method and illustration for calculating the HDI at the household level. This immediately allows the analysis of the HDI by any kind of population subgroups and by household socioeconomic characteristics. Furthermore, it allows to apply any kind of inequality measure to the HDI across population subgroups and over time. We illustrate our approach for 15 developing countries. We find large inequalities within countries between population subgroups, particularly by income, location, and education of the household head. We also find considerable inequality when looking at inequality measures like the Theil or the Gini coefficient; within-group inequality is, however, invariably larger than between-group inequality and inequality in the HDI within countries is of similar order of magnitude of inequality in the HDI between countries.

**Key words:** Human Development Index, Income Inequality, Differential Mortality, Inequality in Education.

---

\*Stephan Klasen (corresponding author) (sklasen@uni-goettingen.de), University of Göttingen, Department of Economics, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany. Phone: +49-551-397303; Fax: +49-551-397302.

Kenneth Harttgen (k.harttgen@wiwi.uni-goettingen.de), University of Göttingen, Department of Economics, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany. Phone: +49-551-398175; Fax: +49-551-397302.

Funding from UNDP in support of this work is gratefully acknowledged.

# 1 Introduction

The HDI is a composite index that measures the average achievement in a country in three basic dimensions of human development: a long and healthy life, measured by life expectancy at birth; education, measured by the adult literacy rate and the gross school enrollment, and standard of living, measured by GDP per capita (UNDP, 2006). Today, the HDI is widely used in academia, the media and in policy circles to measure and compare progress in human development between countries and over time.

Despite its popularity, which is among other things due to its transparency and simplicity, the HDI is criticized for several reasons. First, it neglects several other dimensions of human well-being, such as human rights, security and political participation (see e.g. Anand and Sen (1992), Ranis, Stewart and Samman (2006)). Second, it implies substitution possibilities between the three dimension indices, e.g. a decline in life expectancy can be offset by a rise in GDP per capita<sup>1</sup>. And related to this point, the HDI uses an arbitrary weighting scheme of the three components (see e.g. Kelley (1991), Srinivasan (1994) and Ravallion (1997)).

Perhaps the most constraining weakness is that the HDI only looks at average achievements and, thus, does not take into account the distribution of human development within a country or population subgroup (see e.g., Sagar and Najam (1998)). It is this last issue that we address in this paper.

There are some papers that address the insensitivity of the HDI to inequality between population subgroups. Anand and Sen (1992) and Hicks (1997) suggested to discount each dimension index by one minus the Gini coefficient for that dimension before the arithmetic mean over all three is taken. Therefore, high inequality in one dimension lowers the index value for that dimension and, hence its contribution to the HDI. Although the idea of such a discount factor is rather intuitive, the Gini-corrected HDI has not been widely used, largely due to data constraints.

The gender related development index, or GDI, was another attempt in

---

<sup>1</sup>Moreover, if poor people face higher mortality, their deaths would increase per capita incomes of the survivors, generating a further distortion, particularly in HDI trends over time.

that direction. Its motivation was the 1995 Human Development Report's emphasis on gender inequalities. The GDI adjusts the HDI downward by existing gender inequalities in life-expectancy, education and incomes. The GDI calculates each dimension index separately for men and women and then combines both by taking the harmonic mean, penalizing differences in achievement between men and women. The overall GDI is then calculated by combining the three gender-adjusted dimension indices by taking the arithmetic mean. This concept could of course also be applied using other segmentation variables than gender, such as different ethnic or income groups. This would, however, presume the existence of human development achievement data by groups, which is the topic of our study.<sup>2</sup>

Grimm et al. (2008, 2009) aggregate the three dimensions of the HDI at income quintile levels. Based on a method and computations described in detail in Grimm et al. (2006), the HDR 2006 presented a HDI for all five income quintiles for a sample of 11 OECD countries and 21 developing countries. The results showed that across all countries inequality in human development was very high. It was typically larger in developing countries, and particularly sizable in Africa. This was not only due to an unequal income distribution, but also to substantial inequalities in education and life expectancy. In some middle income developing countries the richest quintile ranked among the high human development countries, whereas the poorest quintile ranked among the low human development countries. But also in rich countries, the differentials were large. Harttgen and Klasen (2009) calculate the HDI separately for internal migrants versus non-migrants. They found small but significant differences in human development between internal migrants and non-migrants. Internal migrants typically show higher outcomes in the HDI than non-migrants.

Another attempt was undertaken by Foster, López-Calva and Székely (2005). They chose an axiomatic approach to derive a distribution sensitive HDI and illustrate this approach for Mexico. They suggest a three-step procedure.

---

<sup>2</sup>However for gender in particular, it is not clear how gender related inequality in income can reasonably be measured. Generally, the GDI uses information on earned income of males and females, based on sex-specific labor force participation rates and earnings differentials (UNDP, 2006). In most cases men and women pool incomes in households. Usually not much information is available how the pooled income is then allocated among household members. That and other critical issues related to the GDI are discussed in detail by Klasen (2006).

First, each dimension index is calculated on the lowest possible aggregation level, given data availability, for instance, income at the level of households and life-expectancy at the level of municipalities (taken from census data). Second, for each dimension an overall index is computed by taking the generalized mean. Third, the overall HDI is computed by taking again the generalized mean instead of the simple arithmetic mean. The advantage of this approach is its axiomatic foundation, for example its decomposability by subgroups. However, the life expectancy index is aggregated at the municipalities level which suppress variation in that sub-index. Furthermore, regarding the enrolment index the analysis is restricted to households with children resulting in a loss of data.

In short, all previous attempts to capture inequality in the HDI have also used aggregate information and there exists no HDI at the household level. This paper provides a method and an illustration for calculating the HDI at the household level. This will allow a large range of previously unavailable analysis to yield new insights with respect to levels and changes of human development. It immediately allows comparisons across population subgroups (e.g. urban, rural), by income and other population groups like the mentioned papers. Furthermore, it provides a completely new opportunity to analyze differences in the HDI between household specific characteristics.<sup>3</sup> In addition, having calculated an HDI at the household level, one could calculate any kind of inequality measure of the HDI, compare it across space and time and decompose it within and between groups.

When constructing distribution-sensitive measures of human development, limited data availability on the distribution of human development achievements seriously constrains the analysis. Today household income surveys are widely undertaken and provide data on income distribution. However, it is much more difficult to get data on inequality in life expectancy, educational achievements and literacy. Thus, the main challenge of calculating a household based HDI is to overcome the data constraints which we face using household

---

<sup>3</sup>Although the HDI will be calculated at the household level, we can extend this analysis to the person-level by imputing the HDI of a household to each member. Of course this would ignore intra-household inequality in the HDI which is quite hard to tackle give our approach.

survey data. First, there is virtually no survey that includes information on income, education and mortality simultaneously. Second, life expectancy is difficult to estimate at the household level and one has to rely on some sort of imputation or simulation techniques. Third, no information on enrolment exists for households without children.

The objective of this paper is first of all illustrative. We will also show that our methodology has some shortcomings. Hence all presented results should be interpreted with caution and in the light of our assumptions. The reminder of this paper is organized as follows. Section 2 presents our methodology. Section 3 presents the sample of countries for which we illustrate it and presents the results. Section 4 concludes.

## **2 Methodology**

### **2.1 Calculating the GDP index**

For our analysis we rely on DHS data where information on education and mortality is available. We start with the calculation of the GDP component of the HDI. Since we do not have information on income or expenditure in the DHS data sets that can be used for our analysis, we consider an alternative approach to determine the socio-economic status of a household, which we use as a proxy for income or expenditure. In particular, we combine an asset index approach defining well-being proposed by Filmer and Pritchett (2001) and Sahn and Stifel (2001) with an income simulation approach proposed by Harttgen and Vollmer (2009). We thereby simulate income levels for each household in the DHS data sets to overcome the problem that the DHS do not contain information on income or expenditure .

We proceed as follows. In a first step, we calculate an asset index( see Filmer and Pritchett, 2001; Sahn and Stifel, 2001). The main idea of this approach is to construct an aggregated uni-dimensional index over the range of different dichotomous variables of household assets capturing housing durables and information on the housing quality that indicate the material status (welfare) of the household:

$$A_i = \hat{\gamma}_1 a_{i1} + \dots + \hat{\gamma}_n a_{in} \quad (1)$$

where  $A_i$  is the asset index, the  $a_{in}$ 's refer to the respective asset of the household  $i$  recorded as dichotomous variables in the DHS data sets and the  $\hat{\gamma}$  are the respective weights for each asset that are to be estimated.

For the estimation of the weights and for the aggregation of the index, we use a principal component analysis proposed by Filmer and Pritchett (2001), relying on the first principal component as our asset index.<sup>4</sup> In particular, as components for the asset index we include dichotomous variables whether the following assets in a household exist or not: radio, TV, refrigerator, bike, motorized transport, capturing household durables and type of floor material, type of wall material, type of toilet, and type drinking water capturing the housing quality and we calculate the asset indices separately for each country and period.<sup>5</sup>

A large body of literature exists using an asset index to explain inequalities in educational outcomes (e.g. Ainsworth and Filmer, 2006; Bicego et al, 2003), health outcomes (e.g. Bollen et al., 2002; Schellenberg et al, 2003), child malnutrition (e.g. Sahn and Stifel, 2003; Tarozzi and Mahajan, 2005), child mortality (e.g. Sastry 2004) when data on income or expenditure is missing. In addition, asset indices are used to analyze changes and determinants of poverty (Harttgen and Misselhorn, 2007; Sahn and Stifel, 2000; Stifel and Christiaensen 2007; World Bank, 2006).

The use of the asset index approach to derive a welfare distribution faces some critical issues that should be mentioned when using this approach. First, the asset index might not correctly reveal differences between urban and rural areas. The asset index can be biased due to usually huge differences in prices and the supply of such assets as well as differences in preferences for assets

---

<sup>4</sup>An alternative way to estimate the weights for the assets to derive the aggregated index is a factor analysis employed, for example, by Sahn and Stifel (2001). However, the two estimation methods show very similar results.

<sup>5</sup>The asset index is calculated for each individual, weighted by the household size. By also using DHS data, Houweling et al. (2003) analyze how the choice of indicators to be included in the asset index leads makes a difference in the ranking of households. The authors find significant but very small differences in the rankings of households depending on different sets of indicators.

between both areas. For example, urban households own other assets than rural households.

Second, the main critical issue of using the asset index is whether it can serve as an appropriate proxy for income or expenditure. Another strand of literature validates the use of an asset index as a proxy of welfare when data on income or expenditure are not available. For example, Stewart and Simelane (2005) validate the use of the asset index as a proxy for income to predict child mortality in South Africa. They find a very close relationship between income and the asset index. The recent paper by Filmer and Scott (2008) provides an excellent validation of the use of various asset index methods by comparing how asset index outcomes match to results using per capita expenditures with respect to the ranking of households and with respect to inequality analysis outcomes in education, health care use, fertility, and child mortality. They show that inferences about inequalities in education and health are robust to the use of the asset index. The gradient of the outcomes of the asset index closely follows the outcome using per capita expenditures. However, although they do find an overlap, they also show some differences in the ranking of households between the asset index and per capita expenditures in the poorest population quintile. The reason for the differences in the ranking of households results is that asset indices are less suitable to capture the existence of transitory shocks, because assets are a measure of stocks whereas income or expenditure are flow measures. In addition, assets indices are typically derived from public goods at the household level, while expenditures prominently captures the consumption of food.

Filmer and Scott (2008) argue that targeting of social program to the poorest population quintile on the basis of the asset index would therefor only partly reach the same households. They found that the assets index identifies especially the more rural and smaller households as deprived, compared to per capita expenditure. They conclude that because the gradient of the economic status is similar to that of per capita expenditure using the asset index would not lead to a misleading targeting and that using the asset index even allows to identify the most deprived households in terms education, health, and labor force participation.



Thus, when relying on the asset index to the welfare rankings based on an asset index do not lead to the exact welfare ranking based on per capita expenditure but the gradient of both measures are similar. Similar results were also found by Harttgen and Volmer (2010) who validate the use of the asset index as a proxy for per capita income using LSMS data for several developing countries and compare the household ranking and outcomes of social indicators of human development. They also find some differences in the ranking of households while also here the gradient of the asset index and household per capita income are similar.

The finding that the asset index even allows to identify the most deprived households reflects the argument that assets may be a better proxy for long-term income than annual income. One advantage of using the asset index as an indicator for the long term capacity of households to purchase goods and services and to cope with different kinds of negative shocks is that the asset index is less vulnerable to fluctuations over time than income or expenditure. For the analysis this means that the welfare ordering of households over time is more consistent compared to the ordering of households based on income. Therefore, using the asset index provides a good indicator of long term well-being, which is in line with the basic idea of the HDI. And the scaling of the assets values with corresponding income values based on the GINI makes our results comparable across time and space and across related studies which examine income inequality.

In a second step, we derive a log normal distribution (LN) based on the respective country specific mean income per capita and the respective Gini coefficient obtained from PovcalNet. Formally, the log-normal distribution  $LN(\mu, \sigma)$  is defined as the distribution of the random variable  $Y = \exp(X)$ , where  $X \sim N(\mu, \sigma)$  has a normal distribution with mean  $\mu$  and standard deviation  $\sigma$ . It can be shown that the density of  $LN(\mu, \sigma)$  is

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-(\log(x)-\mu)^2/2\sigma^2}, \quad x > 0, \quad (2)$$

and its mean and variance are given respectively by

$$E(Y) = e^{\mu+\sigma^2/2}, \quad Var(Y) = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}. \quad (3)$$

We should briefly discuss the interpretation of the parameters  $\mu$  and  $\sigma$ , which is different from that of the normal distribution. In fact, from (3) one sees that  $e^\mu$  is proportional to the expectation and  $(e^\mu)^2$  is proportional to the variance, and in fact,  $e^\mu$  is the scale parameter of the log-normal distribution, whereas  $\sigma$  is a shape parameter. Since the Gini coefficient is invariant under changes of scale (it does not matter whether income is measured in Euro or in Dollar), it should be independent of  $\mu$  and only depend on  $\sigma$ . This is indeed the case: The Gini coefficient  $G$  of  $LN(\mu, \sigma)$  is given by  $G = 2\Phi(\sigma/\sqrt{2}) - 1$ , where  $\Phi$  is the distribution function of the standard normal distribution. Therefore, the parameters  $\mu$  and  $\sigma$  of  $LN(\mu, \sigma)$  can be determined from the average income  $E(Y)$  and the Gini coefficient  $G$  as follows.

$$\sigma = \sqrt{2} \Phi^{-1} \left( \frac{G+1}{2} \right), \quad \mu = \log(E(Y)) - \sigma^2/2.$$

Hence, with the two parameters, the mean income per capita and the Gini, we are able to estimate  $\mu$  and  $\sigma$  of the density function of the log normal distribution for each country.

In a third step, the asset index distribution will also be modeled by a log-normal distribution.<sup>6</sup> In doing so, we now have two log normal distributions, one from the asset index and one national income distribution based on the country specific mean income and the country specific Gini coefficient.

In a fourth step, we can then simulate household income per capita based on the asset index distribution. In particular, we can attach to each quantum of the asset index distribution the respective income value from the income distribution and derive to each asset index value the respective simulated income value. To illustrate this approach, Figure 1 shows the asset index distribution exemplarily for Nicaragua (right) and Burkina Faso and also the obtained income distribution following our approach (left). We can see that the assumption that the asset index follows a log normal distribution holds and that the estimated income distribution closely follows that of the asset index distribution.

[please insert Figure 1 here]

---

<sup>6</sup>The estimation of the distribution is based on a maximum likelihood estimation technique.

Then, in a fifth step, we can easily calculate the household specific GDP component of the HDI. To eliminate differences in price levels across countries we express household income per capita  $y_h$  calculated from the HIS, in USD PPP using the conversion factors based on price data from the latest International Comparison Program surveys provided by the World Bank (2005):

$$y_h^{PPP} = y_h \times PPP. \quad (4)$$

Then, we rescale  $y_h^{PPP}$  using the ratio between  $\bar{y}^{PPP}$  and GDP per capita expressed in PPP (taken from the general HDI):

$$ry_h^{PPP} = y_h^{PPP} \times \left[ \frac{GDPPC^{PPP}}{\bar{y}^{PPP}} \right].^7 \quad (5)$$

Once these adjustments are done, it is straightforward to calculate the household specific GDP index, using the usual minimum and maximum values of the HDI:

$$Y^h = \frac{\log \bar{y}^{h,PPP} - \log(100)}{\log(40,000) - \log(100)} \quad \forall h = 1, 2, \dots, K, \quad (6)$$

where  $\bar{y}^{h,PPP}$  is the household specific arithmetic mean of the rescaled household income per capita.

It should be noted that in richer countries the GDP per capita measure for the richest households could easily exceed 40,000 USD PPP and, hence, the index could take a value greater than 1.<sup>8</sup>

There are two ways to deal with this issue. The first is capping income to the maximum of 40,000 USDPPP. Or second, one does not cap the incomes but rather caps the GDP index to one. To avoid the right-truncation of the income distribution which is needed for calculation of distribution, we suggest do not cap the household's income, since we are also interested to the analysis of inequality in human development.<sup>9</sup> Capping the income would

---

<sup>7</sup>Obviously, it is rather unlikely that the income sources not captured by the household survey are distributed in the same way as the observed income sources (see e.g. Ravallion, 2003). However, we think it would be very difficult to come up with any reasonable alternative rule to correct this bias across all countries without at the same time complicating extensively our methodology.

<sup>8</sup>In the last Human Development Report (UNDP, 2009) such index numbers are set to 1. In this study we do not follow this rule.

<sup>9</sup>This means that the scaling of income to match the country GDP will be done using the uncapped income.

lead to an artificial decrease in potential inequality. The same argument of holds also for the log transformation of income in the GDP component of the HDI. However, to be in line with the way the HDI is measured we use the log-transformation of income and additionally provide in the results section some sensitivity analysis of how the capping and the log transformation of the income affects the outcome of inequality measures.

## 2.2 Calculating the education index

In the next step, we calculate the education index of the HDI at the household level. For this, we need to calculate rates of adult literacy and gross school enrollment at the household level.

For the adult literacy rates, we can directly use the information on literacy in the DHS. For some DHS, the information on literacy is missing. Here, we define an adult household member as being literate if she has at least five years of schooling completed. The data constraint we face when calculating the education index is that enrolment information is only available for households that have school-age children. The main challenge that arises here is the question of how to compare the value for the education component of households where we just have information on literacy with those where we have information on literacy and enrolment. We provide two possible solutions.

First, we drop the enrolment component and rely only on literacy. Here, no assumptions of replacing missing values have to be made. But, on the other hand, this approach could overstate the education component because literacy rates are often higher than enrolment rates. In addition we would lose one sub-component of the HDI. In principle, one could also simply drop the observations for which we do not have information on enrolment, i.e. the households without children in that particular age range. Simply, deleting the missing values might lead to biased results if the remaining cases are not representative for the entire sample (e.g. Schaefer and Graham 2002). Second, we use an imputation-based approach to fill the missing values of enrolment. Imputation using a regression-based approach involves the employment of a deterministic or stochastic regression method to impute the missing values (Landerman et al. 1997). This means that the missing value is replaced by a

regression predicted score, where one uses the existing values of the respective variables and regress them on a set of covariates. In particular, we regress the enrolment status on a set of household and community characteristics and then we use the obtained coefficients to predict the enrolment rate for all households (on not only for those without children. This means we are not filling any observations but rather imputing household-based enrolment rates for all households.<sup>10</sup>

$$\hat{x}_h = a + X_h b + u_h, \quad \forall h = 1, 2, \dots, K, \quad (7)$$

where  $\hat{x}_h$  is the value of imputed value (enrolment rate) of household  $h$ ,  $X_h$  is a vector of socioeconomic characteristics,  $b$  is the vector of regression coefficients. To account for the error term in the regression (and thus to avoid the unwarranted precision of the point estimate of our imputation), we add a random term  $u_h$  drawn from a normal distribution and where its variance is estimated from the sample (stochastic approach). Without including a random term (deterministic approach), the imputation would likely result in an underestimated variance of the variable.

There is a broad literature both on the review and on the application of imputation (e.g., Graham and Hofer, 2000; Rubin, 2004; Stern and Russel, 2001; Schaefer, 1997; Graham et al. 2003; Schaefer and Graham 2002; Allison 2007). However, there is also critic on mean substitution and regression-based single imputation (e.g. Graham et al (2003); Landerman et al. (1997)). The major shortcoming of this approach (besides depending on the quality of the regression - at the current state of the paper we did not take into account a possible selection bias) is that the variance is still underestimated and thus standard errors and significance test can still be biased. However, in this paper we do not want to use the filled values for an analysis of the determinants of education. Since the proposal here is to impute enrolment rates for descriptive purposes, we think that this approach is a reliable method to obtain education estimates for all households (those with and without children but otherwise

---

<sup>10</sup>Important for the imputation is that the means of the socioeconomic characteristics do not differ very much between the samples of with and without information on enrolment. Although present in our sample of countries, differences are not very large.

equal characteristics).<sup>11</sup>

After obtaining an enrolment rate and literacy rate for each household in the data set, we can calculate the household specific education index of the HDI. We calculate the household specific adult literacy index  $A^h$  and gross school enrolment index  $G^h$  using again the corresponding usual minimum and maximum values employed in the HDI

$$A^h = \frac{a^h - 0}{1 - 0} \quad \forall h = 1, 2, \dots, K, \quad (8)$$

$$G^h = \frac{g^h - 0}{1 - 0} \quad \forall h = 1, 2, \dots, K, \quad (9)$$

where  $a^h$  refers to the household specific adult literacy and  $g^h$  to the imputed household specific gross school enrolment rate. The household specific education index  $E^h$  is calculated using the same weighted average as done with the HDI:

$$E^h = (2/3) \times A^h + (1/3) \times G^h \quad \forall h = 1, 2, \dots, K. \quad (10)$$

In addition, we also calculate the education component of the HDI based on another indicator of educational attainment to deal with the issue that adult literacy may not be a very good indicator of educational attainment, because it does not take into account higher levels of achievements in education. In particular, we introduce the indicator of the mean years of schooling of adults aged 25 and older into the education component by dropping the adult literacy rate and leaving the weights to calculate the education index unchanged. With this way, we can illustrate how the choice of the educational indicator influences the outcome of the education index and of the overall HDI. The main challenge that arises using years of education is to normalize the subindex between 0 and 1, because we need to decide on a minimum and

---

<sup>11</sup>The solution to deal with this issue would be to rely on multiple imputation (Rubin, 1977 and 1987; Schfer, 1997). The idea is to repeat the imputation process, producing multiple "complete" data sets. The values are drawn from the Bayesian posterior distribution of the parameters. Because of the random term, the estimates of the parameters will slightly differ and this variability can then be used to adjust the standard errors upwards (Allison 2007). These analysis results are then combined to one overall analysis resulting in the prediction of the missing values (Wayman 2003). It has been shown that multiple imputation performs favorably (see, e.g. Schaefer and Graham, 2002; Schaefer, 1997; Wayman 2003). Multiple imputation allows to produce estimates that are consistent, efficient, and asymptotically norm when the assumption of missing and random (MAR) is fulfilled (Allison, 2007). However, since multiple imputation is very time consuming we leave this for further research.

maximum amount of years of education. In this paper, we define the minimum years of education to be zero and the maximum to be 16 years of schooling.<sup>12</sup>

### 2.3 Calculating the life expectancy index

To calculate the life expectancy index, we combine information on child mortality with model life tables and use again a regression based approach to calculate mortality rates at the household level. The reason for this imputation is twofold. First, we need to overcome the problem of households without children resulting in a loss of data. Second, we need to obtain an estimate of child mortality that has a more continuous character, because otherwise we would have only limited variation in the data since in most household either none, one or two children died resulting in a household specific mortality rate of 0 (for which no life expectancy is computable), 0.33, 0.66 or 1.

First, as already done in the previous section to obtain school enrollment at the household level, we regress child mortality on a set of basic household and community socioeconomic characteristics using a duration model to control for censored data. Then, we use the prediction of child mortality for all households (and not only on those without children). Again, this means we are not filling any observations but rather imputing household-based child mortality rates for all households. And again, one should be very clear that since we are imputing child mortality to households, the HDI we are calculating for each household is not the 'true' HDI of that household (which is unknowable until we know the actual life expectancy of the household members which we only know for sure once they have all died). But it is the HDI for this 'type' of household (with the particular characteristics that affected the imputation).

Second, after having estimated the household specific mortality rate, we apply the recently provided modified logit life table systems by Murray et al. (2003) to estimate the household specific life expectancy at birth. This model is based on a Brass logit approach:

---

<sup>12</sup>In particular, this yields to  $S^h = \frac{s^h - 0}{16 - 0} \quad \forall h = 1, 2, \dots, K.$ , where  $s^h$  refers to the mean years of schooling per household. Of course the choice of the upper and lower limits of the education to calculate the educational sub-index we have results on the outcome of the index. As already discussed in the previous section, capping the years of education results in a loss of potential inequality. However, in this case, the limits for inequality are inherent in the respective school system and not artificially defined for purposes of calculation.

$$\begin{aligned}
Logit(l_x^h) &= \alpha_h + \beta_h * Logit(l_x^s) + \\
\gamma_x \left[ 1 - \left( \frac{Logit(l_5^h)}{Logit(l_5^s)} \right) \right] &+ \theta_x \left[ 1 - \left( \frac{Logit(l_{60}^h)}{Logit(l_{60}^s)} \right) \right]
\end{aligned} \tag{11}$$

$\forall h = 1, 2, \dots, K.$ , where  $x$  is the age,  $\gamma_x$  and  $\theta_x$  are parameters of the age specific Standard Life Table,  $\alpha_h$  and  $\beta_h$  are country specific parameters, and  $\gamma$  the survival probability from zero to  $x$ , 5, and 60. To any value of  $l_5$ , the corresponding value for the life expectancy at birth  $e_0$  can be estimated through in iterative procedure.

The advantages of the modified logit life tables by Murray et al. (2003) compared, for example to Princeton Model Life Tables (Coale and Demeny, 1983) or the older Ledermann model life tables (Ledermann, 1969), are that they are very flexible and rely on more than 1800 recently available life tables.<sup>13</sup>

Third, after having estimated life expectancy for each household in the DHS data, we can then calculate the household specific life expectancy index of the HDI.

$$H^h = \frac{\hat{e}_0^h - 25}{85 - 25} \quad \forall h = 1, 2, \dots, K., \tag{12}$$

An alternative approach to estimate the life expectancy at birth at the household or individual level is provided by the WHO (2001). In principle, this approach follows the same assumption to estimate the life expectancy. Also here, the modified Brass logit system is used to estimate a whole life table for all countries. Since we have life tables for all countries (which reflects the age-specific life expectancies for one representative household), we can then easily get the age specific life expectancy  $e_x$ , i.e. the expected years to live at any given age in a particular country. By adding this value to the respective age of the household member, we then get a value for  $e_0$  for every person.

---

<sup>13</sup>We also compare the results with the outcome based on the Ledermann life tables and also with the outcome of a sample. In fact, we find a considerable overestimation of life expectancy using the older Ledermann approach, which especially is driven that the older model life tables do not allow to capture any effects of the HIV/AIDS epidemic, particularly in Sub-Saharan African countries.



However, two issues arise when using the WHO (2001) approach. The first problem with the WHO approach is that it calculates only 'one' age-specific life expectancy for each country and thereby precisely ignores the within-country inequality in life chances that we want to explore with the household-based HDI.

The second problem is related to the way the HDI employs life expectancy at birth. In particular, this figure is a synthetic number that is an answer to the following question: If a person was born today and then lived through the age-sex specific mortality rates that currently prevail all at once, how long would the life of the person be? Now this figure is not relevant to any individual for two reasons: a) you obviously cannot live through your entire life in one year and b) anyone who is older than 1 can no longer die from the mortality rates that afflict people younger than they are. So their expected length of life will be larger than life expectancy at birth. Hence, the life expectancy component in the HDI is exactly what we want to measure, which is a snapshot of mortality conditions in a country at a certain point in time as an indicator of current life chances. If one actually calculated the expected lengths of life of those people currently alive, that number would be strongly influenced by the age structure of the population<sup>14</sup>. It would also have the consequence of ignoring high infant mortality rates as one only cares about the surviving infants and calculate their life chances and ignore the ones that just died. Therefore we think the life expectancy component as currently conceived in the HDI is just right and, consequently, the life expectancy component we calculate for the HDI at the household level is also favorable to the WHO approach. It measures current mortality conditions for that (type of) household and the impact this has on life chances for people.

We illustrate the difference between the approaches for two countries. Table A1 shows the outcomes of the estimated life expectancy (based on the regression approach and based on the WHO approach) for Armenia and Bolivia. The difference between the two approaches is larger for Bolivia than for Armenia but both are sizable. This is translated into the life expectancy

---

<sup>14</sup>For example if you have few young people and correspondingly a high share of old people, your expected life lengths would be much higher than in a country with many young people; but this is due to mortality conditions of the past, not the present

index, which is for Bolivia 0.68 based on the regression and 0.79 for the WHO approach. What is very interesting that the standard deviation for the WHO approach is very low. This is because the minimum life expectancy is already at a very high level (69), whereas we get lower values for the regression based approach<sup>15</sup>. Hence, the variation in the life expectancy index is relatively low compared to the regression based approach. This has also consequences for an inequality analysis. In fact, the WHO approach reduces possible inequality. Based on UN mortality statistics, Hicks (1997) provides Gini coefficients for life expectancy for 20 countries. The Gini coefficients are higher than those we found in our samples which is due to the fact that Hicks considers data on actual life lengths from 1983-1991 (and thus largely reflecting mortality conditions of people born in the 1930s to the 1980s) and that he (implicitly) imputes a life different expectancy value to all household members while we calculate an average life expectancy for all household members<sup>16</sup>. However, if we would have used the WHO approach to estimate life expectancy, the Gini coefficient would have been even smaller.<sup>17</sup>

## 2.4 Calculating the household-based HDI

Once the three dimension indices are calculated, we simply calculate the household specific HDI, by taking the arithmetic mean of the three dimension indices. We use  $\mu(y)$  to denote the arithmetic mean<sup>18</sup> of a given distribution  $y$ , i.e. household income per capita, and apply this definition also to the education ( $e$ ) and health ( $h$ ) component of the HDI. All three dimensions of the HDI can be represented in a  $3 \times k$  matrix  $D$ , where the first row is the vector  $y$ , followed by  $e$  and  $h$ . The household based human development index  $H$  (where  $k$  refers to the number of households in the data) can then be defined as a function  $F : D \rightarrow R$  from the set of  $D$  matrices to the real numbers  $R$  and

<sup>15</sup>In particular, we capped the values below 25 to 25.

<sup>16</sup>This is not so much an issue of accounting for intra-household inequality in life chances but more of a question of whether and how to adequately account for stochastic inequality in life chances. For example, a 5 person household with an average life expectancy of 50 will likely have some people who die young and others who die much later. We are currently investigating whether there are plausible ways of incorporating this stochastic inequality in life expectancy

<sup>17</sup>For example, whereas for the estimated life expectancy for Armenia, the Gini is 0.15, it is only 0.02 when applying the WHO approach.

<sup>18</sup>the formula for the arithmetic mean is  $\mu(y) = (y_1 + y_2 + \dots + y_k)/k$ .

formally expressed as the mean of the means:<sup>19</sup>

$$H(D)_{household} = \mu[\mu(y), \mu(e), \mu(h)], \quad (13)$$

which corresponds to the mean achievement in each dimension of the HDI which is then averaged across dimensions. To get person-based values, this value is assigned to each household member.

In addition to the traditional HDI, we also apply two inequality adjusted HDI proposed by Foster et al. (2005) and Seth (2009). In particular, the authors extend the traditional HDI by an inequality measure to take into account the distribution of the three dimensions within a population. The Foster et al. (2005) approach is based on the idea to use a general mean of the arithmetic mean to average each dimension of the HDI, namely  $\mu_\beta(y)$ ,  $\mu_\beta(e)$ , and  $\mu_\beta(h)$ , where  $\beta \neq 0$ .<sup>20</sup> General means are sensitive to the distribution in the sense we introduce an inequality aversion parameter  $\beta$ . A  $\beta$  less than zero gives a greater weight to the achievements of the lower end of the distribution, i.e. the poorer households. The higher the inequality, the higher is the importance of the achievements of the poor. For  $\beta = 1$ , the general mean is the arithmetic mean, which is indifferent to inequality. In particular, Foster et al. (2005) extend the Atkinson class of inequality measures (Atkinson, 1970) to multinational HDI.<sup>21</sup> Hence, for each dimension an overall index is computed by taking the generalized mean  $\mu_\beta$ :

$$H_\beta(D)_{household} = \mu_\beta[\mu_\beta(y), \mu_\beta(e), \mu_\beta(h)], \quad \text{for } \beta \neq 0, \quad (14)$$

which we in the following define as FLS. For  $\beta = 1$ ,  $\mu$  yields the arithmetic mean, but for negative values for  $\beta$ ,  $\mu$  gives more emphasis on the lower end of the distribution of each dimension. Now the HDI is expressed as a general mean of the general means. In doing so, the results of IHDI is comparable to the outcomes of the traditional HDI.

<sup>19</sup>See Foster et al. 2005.

<sup>20</sup>The formula for the general means is  $\mu_\beta(y) = [(y_1^\beta + y_2^\beta + \dots + y_k^\beta)/k]^{1/\beta}$ .

<sup>21</sup>The formula for the Atkinson family of inequality measures is  $I_{1-\epsilon}(y) = 1 - [\mu_{1-\epsilon}(y)/\mu(y)]$  for  $\epsilon > 0$ . This means, the Atkinson inequality measure subtracts one minus the ratio of the general mean and the arithmetic mean, where  $\epsilon$  can be interpreted as an inequality aversion parameter ( $\beta = 1 - \epsilon$ ). For  $\beta = 1 \rightarrow \epsilon = 0$ , the general mean is the arithmetic mean. Greater inequality is reflected in a higher ratio between the general 'distribution sensitive' mean and the 'neutral' arithmetic mean.

In addition to the FLS measure, we apply another distribution sensitive HDI proposed by Seth (2009). Also here, the measure uses a proximate Atkinson measure of inequality to adjust the traditional HDI. In particular, each dimension index is multiplied by  $(1-A)$ , where  $A$  refers to the Atkinson inequality measure. For both the Seth measure and the FLS measure we provide results for several values of the inequality aversion parameter  $\epsilon$  ( $\epsilon = 0, \epsilon = 0.5, \epsilon = 2$ , and  $\epsilon = 3$ ).

### 3 Results

#### 3.1 Overall results

In this section, we present the results of the household-based HDI for our 15 countries. Table 1 shows the mean household-based HDI and its sub-components by country and also the outcomes for different approaches to calculate the household-based education index.<sup>22</sup> HDI 1 refers to the approach, where we simply drop the enrolment component and only rely on adult literacy, HDI 2 refers to the regression based approach to impute literacy and enrolment, and HDI 3 refers to the approach where we use the imputed gross school enrolment and years of education as the indicator of educational attainment.

With respect to the different approaches to calculate the education index, the differences are shown in the last three columns of Table 1. We see small but significant differences between the regression based approach to impute literacy and enrolment and simply using the adult literacy rate to calculate the education index. Relying only on literacy and, thus, taking only one indicator of educational attainment into account, we potentially either underestimate or overestimate the education component compared to the approach where enrollment is also used, because the adult literacy rate is often either considerably lower or higher than the enrolment rates. This is illustrated in Table 1, which shows the descriptive statistics for all indicators. For example, in Armenia and Bolivia, the differences between enrolment and literacy are very high, which is translated into a much higher value for the education index relying only on literacy. Conversely, in the poorest African countries, including

---

<sup>22</sup>For all the results presented in this section, we do not provide any confidence intervals or significance tests between differences in the outcomes because of space limitations. This will be provided in the next version of the paper.

enrolment ratios leads to higher HDIs as they are higher than literacy levels. We find even larger level differences in the education index when we use years of education as the indicator of educational outcome instead of literacy. The education index based on years of schooling of adults aged 25 and older shows much lower outcomes than the other two approaches (see the last column of Table 1). This is because the mean values are considerably lower than the maximum of assumed 16 years of education achievable (see Table 1). These differences are then translated into an overall HDI which is significantly lower. Figures 2 and 3 provide the differences in the distribution between the alternative education indices and the alternative HDI outcomes. This has an important implication considering a possible change in the calculation of the HDI for future Human Development Reports. The main question that arises here, is how one would compare the results of previous reports, because the values of the HDI are expected to be much higher. This would lead to a misleading interpretation of a decline in outcomes of human development.

However, besides differences in the level of the education index, the alternative approaches to calculate the education index have no impact on the ranking of countries. Regardless of what approach is chosen, the ranking between countries of the total HDI remains unchanged. This means for example, that Burkina Faso remains the country with the lowest value whereas Armenia remains the country with the highest outcome of the HDI.<sup>23</sup>

The results of the estimated household-based HDI have to be treated with caution also in the sense that the imputing, which is based partly on the same characteristics, can lead to an in-built correlation for health and education due to common covariates in the regression. However, the same argument would be true for the three components of the HDI in general. To investigate this issue, we provide in Table A3 in the Appendix the correlation coefficients between the indicators that enter the index. We see that although there is a correlation, the correlation coefficient between indicators are not very high, leaving enough scope for heterogeneity between the three dimensions.

---

<sup>23</sup>The rankings provided in Table 1 show the respective rank position of the countries for the particular year in question. This means, that although Burkina Faso shows a lower value for the HDI it has a higher rank position than Ethiopia, because the position is based on different years (2003 and 2005).

Regardless of the educational indicator, Table 1 reveals that Armenia shows the highest level in human development in our sample of countries with an HDI 2 value of 0.769 followed by Egypt (0.684), whereas the lowest value is found for two African countries, namely Burkina Faso (0.391) and Ethiopia (0.401). The high value of the HDI for Armenia is mainly driven by the high outcome in the life expectancy component (0.853) and the high outcome in the education component (0.837), both are also the highest in the sample. Although the GDP index is also high (0.618) it is not the highest value. Concerning levels in income, Egypt even shows a higher GDP index of 0.632. But since both the education index and the life expectancy index are considerably lower (0.630 and 0.789 respectively), the overall HDI is lower than for Armenia. This nicely illustrates the substitution possibilities between the three sub-components of the HDI. The higher education and life expectancy indices offset the relatively lower level of the GDP index. The same holds for Burkina Faso and Ethiopia, whereas the GDP index for Ethiopia is slightly lower (0.370 compared to 0.368), but Burkina Faso performs considerably lower in terms of education and life expectancy.

With respect to the question of what determines the variations in the overall outcomes, we find that variations in life expectancy outcomes are relatively low compared to the outcomes in education and the GDP component.<sup>24</sup> Whereas the life expectancy index ranges from 0.514 (Nigeria) to 0.853 (Armenia), the GDP index ranges from 0.344 (Nigeria) to 0.632 (Egypt) and the education index ranges from 0.212 (Burkina Faso) to 0.837 (Armenia), which is almost 4 times higher.

Table 2 provides the results for the inequality adjusted HDI measures, the FLS and the Seth measure. Table 2 shows the outcomes for different degrees of inequality aversion (defined by the value of  $\epsilon$ ). We clearly see that as higher the inequality aversion parameter as lower are the outcomes of the inequality adjusted HDI, which holds especially for countries with overall lower levels of human development. This is also shown when looking at Figure 2, which shows the distribution of the HDI and the FLS measure. While these measures

---

<sup>24</sup>The same results are observable when looking at the official Human Development Reports.

affect the levels of the HDI, it does not much affect the rankings between our countries (although the ratio of the inequality-adjusted to unadjusted HDIs does differ between countries).

We now turn to the analysis of outcomes of the HDI by different population subgroups and by household characteristics as well as to an analysis of inequality in human development. For all other results in this section, we focus on the use of the regression based approach to estimate literacy and enrolment (HDI 2).

### **3.2 Results by population subgroups and household characteristics**

In this subsection we provide the results of the outcome of the household-based HDI by different population subgroups and household characteristics. Table 3-7 present the HDI by HDI quintiles, by income quintiles, by rural and urban areas, by household size, by the level of education of the household head, by the sex of the household head, and by the age of the household head. The respective tables for the subcomponents are found in the in the Appendix (Tables A3-A18).

Table 3 decompose the outcomes in human development by HDI quintiles itself. This provides us with a first sense of inequality in the outcome of human development. Table 3 shows large inequalities between the poorest and the richest HDI quintile within countries. For example, in Nigeria the ratio of the richest to the poorest quintile is 3.263. The results of Table 3 suggest that inequality tend to be higher in settings where the level of human development is relatively low. The lower the values of the HDI (Table 1, the higher are the differences between the poorest and the richest HDI quintile. This holds also when looking at the distribution of the HDI by income quintiles, which is shown in Table 4. Also here, we observe a large inequality between the poorest and the richest income quintile and also that this inequality is associated with lower levels of human development. Of course, the results for the income quintiles are not unexpected as the income component is inherent in the HDI. But this clear distributional pattern is also observed when the life expectancy index and especially the education index is analyzed by income quintiles (see

Table A6-A8). In particular, we find the largest inequality between the poorest and richest income quintile in the education component <sup>25</sup> Similar results for the outcomes of the HDI and its subcomponent by income quintiles were also found in previous studies by Grimm et al. (2008, 2009).

Table 5 shows the HDI by urban and rural areas. Also here, we find a clear trend. As expected, rural areas are worse off than urban areas with respect to human development. The differences are not as large compared to income quintiles but they are always sizable. For example, in Nicaragua, the ratio between rural and urban areas in the HDI is 0.721. The differences tend to be larger in poorer countries, particularly in Africa and are smallest in Armenia, again driven to an important extent by low differentials in education and health there. And again, similar findings are also found for the sub-components of the HDI.<sup>26</sup> The same differences are also found when looking at the alternative inequality adjusted HDI measure. In particular, Table A10 shows the results for the FLS measure separately by urban and rural areas. We find that once a higher inequality aversion is introduced, the ratio between rural and urban outcomes also rises.

The same clear distributional pattern is found for education of the household head. Households, where the head has no education are considerably worse off than better educated households (Table 6. For example, Zambia shows a HDI that is almost twice as high for households where the household head has achieved higher education compared to households where the head has no educational attainment at all (0.344 compared to 0.629). Again, the differentials are particularly large in Africa. A similar pattern, but to lesser extent is found when looking at the outcomes in the HDI by the age of the household head. Although the inequality, is much lower than for other household characteristics, households with older household heads experience, on average, a higher HDI than households with younger household heads.<sup>27</sup>

<sup>25</sup>For example, in Burkina Faso the richest income quintile show an education component that is 4.5 times higher than the poorest quintile (Table A7).

<sup>26</sup>See Table A9.

<sup>27</sup>However, these results should be treated with caution, because they are also driven by differences in the shares of households of the respective age ranges and thus the calculation are based in very different numbers of observation. For example, there are many more households with a household head aged between 20 and 29 than aged 60 years or older. See also Table A16-A18 for the results for the components of the HDI by the age of the household head.



Quite surprisingly, no clear distributional pattern is found between male and female headed households (Tables 8). First, the differences are not very large, and, second, for some countries outcomes are higher for female headed households than for male headed households (e.g. Egypt) whereas the opposite is found for other countries (e.g. Ethiopia). One explanation would be that female headed households tend to be smaller than male headed households, which would explain the sometimes higher value for the HDI and its sub-components (see also and Table A16). Also for different household sizes no clear distributional pattern in the outcome of the HDI is found (Table 9). In some countries, smaller households show higher HDI outcomes than larger households, in some countries again the opposite finds is found. However, in 10 from 15 countries larger households (more than 11 household members) show a lower HDI than smaller households (size 1-5).<sup>28</sup>

To summarize the foregoing results, we identified significant differences between three alternatives ways to calculate education index. We found large differences in human development across HDI quintiles and income quintiles. The richest population quintile shows much higher outcomes in human development than the poorest population quintile with respect to all three sub-components of the HDI, whereat the largest differences are found for the education component. Furthermore, we found that human development in urban areas is considerably higher than in rural areas, revealing substantial differences in Africa. We also find that the age and education of the household head matters. Older households and households where the head has higher education achieve higher outcomes in the HDI. However, no clear picture for headship and household size (i.e. only small differentials).

### 3.3 Inequality Measures and Decompositions

In addition to the household specific HDI, we also calculated two inequality measures. In particular, we calculated the Gini coefficient for the HDI and its subcomponents. In addition , we provide also the Theil index for the HDI and decompose the measure by within and between inequality for several household characteristics.

---

<sup>28</sup>See Table A18 for the results of the sub-components.

Table 10 shows the Gini coefficient by countries for the HDI and its sub-components. Although it is hard to interpret the absolute value of the Gini (see also below), we can compare the outcome across countries and groups. Table 10 shows that higher values of inequality is found for those countries whose already have shown low levels of human development. For example, Burkina Faso is the country with the highest Gini in the HDI (0.21) and at the same time is shows the lowest value of the HDI in our sample (see Table 1). On the other hand, Armenia (0.06) has the lowest value of the Gini coefficient for the HDI while at the same time it shows the second highest value of the HDI (see Table 1).

Why are the Gini coefficients relatively low compared to usual income Gini coefficients? Overall, the Gini coefficients for the HDI are considerably lower compared to the typical findings for income Gini coefficients. The reason for this relatively low inequality outcome is twofold. First, the main contribution to this low value is driven by the low level of inequality in the GDP index. The low values of the Gini coefficient for the GDP index nicely illustrates how the log transformation of the GDP component reduces inequality. Table 10 provides also the Gini coefficient for the income, the GDP index without the log transformation of income and for the GDP index where the incomes were capped to the value of 40000. We can see that the Gini coefficients for the household per capita income show the expected values that nearly correspond to the official values of the countries taken from PovcalNet.<sup>29</sup> The same holds for the GDP index without the log transformation and for the GDP index based on the capped household income per capita.<sup>30</sup> This means, once we do the log transformation of the income component, we reduce artificially the potential inequality. This means, by using the log transformation, we face a trade-off between taking into account the diminishing rates of return of higher

---

<sup>29</sup>The reason for these small differences is that the asset index distribution is less continuous than the income distribution. This means, for the imputation of the household per capita income we do not take the whole income distribution, but rather draw from the distribution for the values of the asset index distribution.

<sup>30</sup>There is virtually no difference between the GDP index based in the capped and the uncapped income in our sample, because all these countries are relatively poor countries compared to OECD countries for which some countries like Norway exceeds a value of 40,000. In our case, only very few household show higher income values than the threshold resulting an similar values of the GDP index.

income on human development on the one hand and the focus of assessing the degree of inequality within a country or population subgroup on the other.

The second reason for relatively lower Gini coefficients in the HDI stems from the quite low Gini coefficients in the education index and the life expectancy index. We find considerably lower levels in education in the social dimension compared to the income dimension of human development. This mainly stems from the discrete character of the education indicators, meaning that the potential for inequality is lower due to their boundaries of, for example, 0 and 16 years of schooling.<sup>31</sup> This is a kind of inherent capping of the possible values. Some authors (most prominently Thomas et al. (2001, and 2002)) have tried to deal with the issue and have also calculated educational Gini coefficients in education to measure educational inequality based on discrete indicators of educational attainment. Thomas et al (2002) provide the Gini coefficient and Theil indices for years of schooling between 1960 and 2000. The results look quite similar to what we found. The same argument holds for inequalities in life expectancies. We find quite lower inequality in life expectancy than in income. Unfortunately, only very limited comparable Gini coefficients on health exist in the empirical literature. One exception is a study by the Pan American Health Organization (PAHO) (2001) which calculates Gini coefficients for infant mortality for five countries from Latin America. Also here, the results are quite similar to our results in a sense that Gini coefficients for the health indicators are lower than for income indicators.

We also provide the Theil measure and a within and between subgroups decomposition, in particular for income quintiles, rural and urban areas and by education of household head (Table 11). Also here we found relatively low levels of inequality for the countries in the sample. For all subgroups we found that within-group inequality is larger than between-group inequality; this shows that the heterogeneity within groups is a more important driver of human development than the differential between groups (a finding that is usually also found for most groups when income inequality is decomposed into between and within group terms).

---

<sup>31</sup>Although income is also a discrete variable, it has a much more continuous character than, for example, years of schooling.

Another way to interpret our findings on inequality within in the HDI in countries is to compare it to inequality in the HDI between countries. In the literature on income inequality, we observe that income inequality between countries tends to be larger (with Gini coefficients of 0.5-0.8 depending on whether income is PPP adjusted or not) than within-country income inequality in most countries. We therefore examine the same relation now by calculating the Gini coefficients for the HDI and its subcomponents between countries (for the year 2004) taken from the Human Development Report 2006. The Gini coefficient for the overall HDI between countries is 0.14, Gini coefficient for the life expectancy index is 0.16, the Gini coefficient for the education index is 0.12, and the Gini coefficient for the GDP index is 0.16. These results show that the Gini coefficient also between countries is much lower for the HDI than for per capita income levels between countries. Also, we find that inequality between countries in the HDI is larger than inequality in the HDI within most countries; only in a few African countries is inequality in the HDI larger than between countries.

## 4 Conclusion

This paper provides a method and illustration for calculating the HDI at the household level. A household-based HDI provides us we a large range of previously unavailable types of analysis. On the one hand, it immediately allows the analysis of the HDI by any kind of population subgroups and by household socioeconomic characteristics. On the other hand, it allows to apply any kind of inequality measure to the HDI across across population subgroups and over time.

The results of our empirical illustration for 15 developing countries provide new insights with respect to differences in the levels and inequality in human development by population subgroups. We found large inequalities within countries between population subgroups. We found large differences in human development across HDI quintiles and income quintiles. The richest population quintile shows much higher outcomes in human development than the poorest population quintile with respect to all three sub-components of the HDI. Furthermore, we found that human development in urban areas is

considerably higher than in rural areas, revealing substantial differences in Africa. We also found that the age and education of the household head matters. Older households and households where the head has higher education achieve higher outcomes in the HDI. However, no clear picture is found for the household headship and household size, for which we only found minor differences. We also find considerable inequality when looking at inequality measures like the Theil or the Gini coefficient. First, the Gini within countries in social dimensions of human development are lower than for the income dimension but still sizable, Second, countries with lower levels in human development also show higher outcomes in inequality. Third, within population subgroup inequality is larger than between group inequality.

The main challenge of calculating a household-based HDI is due to data limitation. We solve this problems using various kind of imputation techniques to estimate the three subcomponents of the HDI, which rely to some extent on strong methodological assumption. However, these strong assumptions can be justified by applying reasonable approaches to overcome data problems. And despite its methodological shortcomings, this approach hopefully enhances the discussion of measurement issues concerning the HDI.

## References

- Ainsworth, M. and D. Filmer (2006), Children's Schooling: AIDS, Orphanhood, Poverty, and Gender, *World Development* 34(6):1099-1128.
- Allison, P. D. (2007), Missing Data, Quantitative application in social sciences No. 07-136, Sage university press.
- Anand S. and A. Sen (1992), Human Development Index: Methodology and Measurement. Human Development Report Office Occasional Paper 12, UNDP, New York.
- Atkinson, A.B. (1970), On the measurement of inequality, *Journal of Economic Theory*, 2 (3): 244-263.
- Bicego, G., S. Rutstein, and K. Johnson (2003), Dimensions of the Emerging Orphan Crisis in Sub-Saharan Africa, *Social Science and Medicine* 56(6):1235-1247.
- Bollen, K. A., J. L. Glanville, and G. Stecklov (2002), Economic Status Proxies in Studies of Fertility In Developing Countries: Does the Measure matter?, *Population Studies* 56(1):81-96.
- Coale A.J. and P. Demeny, *Regional model life tables and stable populations*. 2. ed. New York/ London: Academic Press, 1983.
- Foster J.E., L.F. López-Calva und M. Székely, (2005), Measuring the Distribution of Human Development: methodology and an application to Mexico, *Journal of Human Development*, 6 (1).
- Fuentes, R., T. Prtze, and P. Seck (2006), Does Access to Water and Sanitation Affects Child Survival? A Five Country Analysis, Human Development Report Office Occasional Paper 2006/04, UNDP.
- Graham, J. W., Cumsille, P. E., and Elek-Fisk, E. (2003). Methods for handling missing data. In J. A. Schinka and W. F. Velicer (Eds.). *Research Methods in Psychology* (pp. 87-114). Volume 2 of *Handbook of Psychology* (I. B. Weiner, Editor-in-Chief). New York: John Wiley and Sons.

- Graham, J. W., and Hofer, S. M. (2000), Multiple imputation in multivariate research, in T. D. Little, K. U. Schnabel, and J. Baumert, (Eds.), *Modeling longitudinal and multiple-group data: Practical issues, applied approaches, and specific examples*, (201-218). Hillsdale, NJ: Erlbaum.
- Grimm M., K. Harttgen, M. Misselhorn and S. Klasen (2006), A Human Development Index by Income Groups. Ibero-American Institute for Economic Research Discussion Paper No. 155, Ibero-American Institute for Economic Research, University of Gttingen.
- Grimm, M., K. Harttgen, M. Misselhorn, and S. Klasen (2008), A Human Development Index by Income Groups, *World Development*, 36 (12): 2527-2546.
- Grimm M., K. Harttgen, M. Misselhorn, T. Munzi, S. Klasen, and T. Smeeding (2009), Inequality in Human Development: An empirical assessment of thirty-two countries, *Social Indicator Research* DOI 10.1007/s11205-009-9497-7.
- Harttgen, K. and S. Vollmer (2010), Validate the use of an asset index as a proxy for income, University of Goettingen, mimeo.
- Harttgen, K. and M. Misselhorn (2006), A Multilevel Approach to Explain Child Mortality and Undernutrition in South Asia and Sub-Saharan Africa, Ibero-America Institute Discussion Paper No. 152.
- Harttgen, K. and S. Klasen (2009), A Human Development Index by Internal Migrational Status, Human Development Research Paper 2009/54, UNDP, submitted to Journal of Human Development and Capabilities.
- Hicks D.A. (1997), The inequality-adjusted human development index: a constructive proposal, *World Development* 28 (8): 1283-1298.
- Houweling, T.A.J., A. E. Kunst, and J. P. Mackenbach (2003), Measuring Health Inequality among Children in Developing Countries: Does the Choice of the Indicator of Economic Status Matter?, *International Journal for Equity in Health* 2(8).

- Kelley A.C., The Human Development Index: 'Handle with Care', *Population and Development Review*, 17 (2), 1991, pp 315-324.
- Klasen S. (2006), UNDPs Gender-Related Measures: Some Conceptual Problems and Possible Solutions. *Journal of Human Development*, 7 (2): 243-274.
- Landerman, Lawrence R., Kenneth C. Land and Carl F. Pieper (1997), An Empirical Evaluation of the Predictive Mean Matching Method for Imputing Missing Values, *Sociological Methods and Research* 26: 3-33.
- Ledermann S., *Nouvelles tables-types de mortalité*. Travaux et documents, Cahier n. 53, Paris: INED and PUF, 1969.
- Little, R. J. A. and D. B. Rubin (1987), *Statistical Analysis with missing data*, New York: Wiley.
- Murray C.J.L., B.D. Ferguson, A.D. Lopez, M. Guillot, J.A. Salomon, and O. Ahmad (2003). Modified logit life table system: principles, empirical validation, and application, *Population Studies*, 57(2): 165-182.
- Pan American Health Organization (PAHO) (2001), Measuring Health Inequalities: Gini Coefficient and Concentration Index, *Epidemiological Bulletin / PAHO*, Vol. 22, No. 1, PAHO.
- Ranis G., F. Stewart and E. Samman, Human Development: beyond the HDI. QEH Working Paper Series - QEHWPS135, 2006.
- Ravallion M. (2003), Measuring Aggregate Welfare in Developing Countries: How well do National Accounts and Surveys Agree? *Review of Economics and Statistics*, 85 (3), pp 645-652.
- Ravallion M., Good and bad growth: The Human Development Reports. *World Development*, 25 (5), 1997, pp 631-638.
- Rubin, D. B. (1977), The design of a general and flexible system for handling non-response in sample surveys, manuscript prepared for the U.S. Social Security Administration.



- Rubin, D. B. (1987), Multiple Imputation for Non-Response in Surveys, John Wiley and Sons, Inc..
- Sagar A.D. and A. Najam, The Human Development Index: A Critical Review. *Ecological Economics*, 25, 1998, pp 249-264.
- Sahn, D. E. and D. Stifel (2003), Exploring Alternative Measures of Welfare in the Absence of Expenditure Data, *Review of Income and Wealth* 49(4): 463-489.
- Sahn, D. E. and D. Stifel (2000), Poverty Comparisons over Time and Across Countries in Africa, *World Development* 28(12):2123-2155.
- Sastry, N. (2004), Trends in Socioeconomic Inequalities in Mortality in Developing Countries: The case of Child Survival in Sao Paulo, Brazil, *Demography* 41(3): 443- 464.
- Schafer. J. L. and Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7 (2), 147-177.
- Schaefer, J. L. (1997). *Analysis of incomplete multivariate data*, London: Chapman and Hall.
- Schellenberg, J. A., C. G. Victora, A. Mushi, D. de Savigny, D. Schellenberg, H. Mshinda, and J. Bryce (2003), Inequities among the Very Poor: Health Care for Children in Rural Southern Tanzania, *The Lancet* 361(9357): 561- 566.
- Stewart, T. and S. Simelane (2005), Are Assets a Valid Proxy for Income? An Analysis of Socioeconomic Status and Child Mortality in South Africa, Indiana University, mimeo.
- Srinivasan T.N., Human Development: A New Paradigm or Reinvention of the Wheel? *American Economic Review*, 84 (2), 1994, pp 238-243.
- Stifel, D. and L. Christiaensen (2007), Tracking Poverty Over Time in the Absence of Comparable Consumption Data, *The World Bank Economic Review* 21(2):317-341.

- Tarozzi, A. and A. Mahajan (2005), Child Nutrition in India in the Nineties: A Story of Increased Gender Inequality?, Unpublished manuscript. Duke University and Stanford University.
- Thomas, Vinod, Yan Wang, and Xibo Fan. 2001. Measuring Education Inequality: Gini Coefficients of Education. Policy Research Working Paper 2525. World Bank, World Bank Institute, Washington, D.C.
- Thomas, Vinod, Yan Wang, and Xibo Fan (2002), A New Dataset on Inequality in Education: Gini coefficients and Theil indices for 140 Countries, 1960-2000, World Bank, mimeo.
- UNDP, *Human Development Report 2009. Overcoming barriers: Human mobility and development*. UNDP, New York, 2005.
- UNDP, *Human Development Report 2006. Beyond Scarcity: Power, Poverty and the Global Water Crisis*. UNDP, New York.
- UNICEF (2009), The state of the world's children - special edition - Statistical Tables, UNICEF.
- Wayman, J. C. (2003), Multiple Imputation for Missing Data: What is it and How Can We Use it? , paper presented at the 2003 Annual Meeting of the American Educational Research Association, Chicago.
- WHO (2001), National Burden of Disease Studies: A Practical Guide, WHO, Geneva.
- World Bank (2006), *World Development Report 2007: Development and the Next Generation*, The World Bank and Oxford University Press. Washington, DC.
- World Bank (2005), World Development Indicators, CD-ROM. World Bank, Washington D.C..

## Tables and figures

Table 1: Overall HDI and sub-components by country (with ranking)

Country	Year	HDI	HDI 2	HDI 3	Ranking HDI 1	Ranking HDI2	Ranking HDI 3	GDP index	Life index	Education index (only literacy)	Education index (regression approach)	Education index (years of education)
Armenia	2005	0.813	0.769	0.683	57	86	119	0.618	0.853	0.968	0.837	0.579
Burkina Faso	2003	0.371	0.391	0.363	171	168	171	0.370	0.593	0.150	0.212	0.126
Bolivia	2003	0.612	0.582	0.527	126	129	135	0.444	0.681	0.711	0.620	0.451
Egypt	2007	0.700	0.684	0.632	119	127	132	0.632	0.789	0.679	0.630	0.473
Ethiopia	2005	0.372	0.401	0.370	177	176	176	0.368	0.521	0.227	0.314	0.215
India	2005	0.642	0.641	0.584	129	129	132	0.551	0.833	0.543	0.541	0.367
Indonesia	2003	0.704	0.674	0.606	111	117	126	0.564	0.765	0.784	0.694	0.487
Kyrgyz Republic	1997	0.720	0.670	0.608	90	107	124	0.479	0.723	0.957	0.807	0.621
Nicaragua	2000	0.554	0.563	0.516	124	124	134	0.452	0.732	0.477	0.504	0.359
Nigeria	2003	0.464	0.469	0.418	157	157	164	0.344	0.514	0.535	0.550	0.394
Pakistan	2007	0.536	0.529	0.478	149	151	165	0.521	0.635	0.452	0.432	0.277
Peru	2005	0.680	0.661	0.603	110	119	129	0.566	0.707	0.767	0.710	0.532
Senegal	2005	0.424	0.450	0.409	163	160	169	0.443	0.579	0.249	0.327	0.204
Vietnam	2002	0.690	0.670	0.605	111	115	127	0.485	0.833	0.752	0.691	0.497
Zambia	2002	0.506	0.477	0.422	138	150	161	0.314	0.533	0.672	0.584	0.413

*Note:* HDI 1 is based only on literacy; HDI 2 is based on the regression based approach for literacy and enrolment. HDI 3 is based on the regression based approach for enrolment and on years of schooling per household aged +25.

*Source:* Demographic and Health Surveys (DHS); calculations by the authors.

Table 1: Descriptive statistics

Country	Enrol- ment rate	Literacy rate	Years of education	Income	Scaled income	Child mortality	Life expctancy
Armenia (2005)	0.69	0.97	8.76	990	4856	21	77.48
Burkina Faso	0.22	0.15	0.96	556	1174	171	57.94
Bolivia	0.47	0.71	6.80	1977	2510	72	65.85
Egypt (2007)	0.61	0.68	6.67	1310	5192	35	72.55
Ethiopia (2005)	0.37	0.23	1.96	600	1026	143	56.27
India (2005)	0.54	0.54	4.49	594	3160	32	75.74
Indonesia (2003)	0.62	0.78	6.96	588	3371	48	71.15
Kyrgyz R. (1997)	0.60	0.96	10.26	745	2154	59	68.48
Nicaragua (2000)	0.56	0.48	4.19	1485	2312	54	68.99
Nigeria (2003)	0.59	0.54	4.82	481	1075	144	55.84
Pakistan (2007)	0.52	0.45	3.56	831	2638	89	63.11
Peru (2005)	0.71	0.77	7.36	2014	4691	65	67.43
Senegal (2005)	0.39	0.25	1.57	789	1793	110	59.76
Vietnam (2002)	0.69	0.75	6.81	661	2209	26	75.85
Zambia (2002)	0.46	0.67	6.07	504	869	134	56.98

*Note:* Enrolment rate refers to the gross enrolment rate, literacy refers to the literacy rate of adults aged 15+. Years of education refers to the mean years of education of per household of adults aged 25+. Household income per capita refer is expressed in USDPPP. Scaled income refers to the household per capita income that is scaled to the national GDP per capita for the respective country and year taken from the Human Development Report. Child mortality refers to the number of dead children before reaching the age of five per 1000 children

*Source:* Demographic and Health Surveys (DHS); calculations by the authors.

Table 2: Household based HDI, FLS and Seth measure

Country	Year	HDI 2	FLS measure			Seth measure				
			FLS (e=0)	FLS (e=0.5)	FLS (e=2)	FLS (e=3)	Seth (e=0)	Seth (e=0.5)	Seth (e=2)	Seth (e=3)
Armenia	2005	0.769	0.769	0.759	0.741	0.725	0.769	0.766	0.696	0.318
Burkina Faso	2003	0.391	0.391	0.361	0.292	0.264	0.391	0.375	0.207	0.060
Bolivia	2003	0.582	0.582	0.566	0.534	0.514	0.582	0.577	0.414	0.075
Egypt	2007	0.684	0.684	0.676	0.657	0.645	0.684	0.682	0.590	0.193
Ethiopia	2005	0.401	0.401	0.382	0.342	0.322	0.401	0.396	0.245	0.057
India	2005	0.641	0.641	0.625	0.591	0.571	0.641	0.635	0.531	0.243
Indonesia	2003	0.674	0.674	0.665	0.644	0.630	0.674	0.672	0.571	0.152
Kyrgyz Republic	1997	0.670	0.670	0.655	0.618	0.594	0.670	0.662	0.563	0.214
Nicaragua	2000	0.563	0.563	0.548	0.514	0.495	0.563	0.557	0.420	0.136
Nigeria	2003	0.469	0.469	0.448	0.417	0.394	0.469	0.465	0.259	0.043
Pakistan	2007	0.529	0.529	0.514	0.484	0.465	0.529	0.526	0.377	0.089
Peru	2005	0.661	0.661	0.651	0.630	0.616	0.661	0.659	0.539	0.160
Senegal	2005	0.450	0.450	0.430	0.387	0.363	0.450	0.444	0.296	0.106
Vietnam	2002	0.670	0.670	0.654	0.620	0.597	0.670	0.662	0.579	0.321
Zambia	2002	0.477	0.477	0.456	0.416	0.391	0.477	0.469	0.305	0.049

*Note:* HDI 1 is based only on literacy; HDI 2 is based on the regression based approach for literacy and enrolment. HDI 3 is based on the regression based approach for enrolment and on years of schooling per household aged +25.

*Source:* Demographic and Health Surveys (DHS); calculations by the authors.

Table 3: HDI by HDI quintiles

Country	Year	By HDIquintiles					Total	Ratio 5:1
		1	2	3	4	5		
Armenia	2005	0.646	0.742	0.782	0.816	0.867	0.771	1.342
Burkina Faso	2003	0.205	0.318	0.387	0.452	0.597	0.385	2.908
Bolivia	2003	0.350	0.504	0.595	0.675	0.781	0.583	2.228
Egypt	2007	0.509	0.625	0.694	0.756	0.837	0.692	1.644
Ethiopia	2005	0.224	0.319	0.386	0.460	0.596	0.380	2.658
India	2005	0.464	0.586	0.650	0.711	0.789	0.618	1.702
Indonesia	2003	0.499	0.627	0.690	0.744	0.811	0.674	1.625
Kyrgyz Republic	1997	0.523	0.625	0.679	0.726	0.792	0.668	1.513
Nicaragua	2000	0.341	0.483	0.568	0.654	0.771	0.586	2.261
Nigeria	2003	0.210	0.381	0.485	0.579	0.685	0.463	3.263
Pakistan	2007	0.336	0.459	0.536	0.608	0.708	0.530	2.107
Peru	2005	0.456	0.593	0.677	0.748	0.835	0.681	1.833
Senegal	2005	0.256	0.376	0.447	0.524	0.655	0.462	2.559
Vietnam	2002	0.494	0.624	0.691	0.740	0.799	0.671	1.620
Zambia	2002	0.279	0.399	0.476	0.555	0.678	0.489	2.431

*Note:* HDI is based on the regression based approach for literacy and enrolment.

*Source:* Demographic and Health Surveys (DHS); calculations by the authors.

Table 4: HDI by income quintiles

Armenia	2005	0.715	0.746	0.766	0.785	0.828	0.770	1.158
Burkina Faso	2003	0.290	0.327	0.361	0.406	0.571	0.384	1.970
Bolivia	2003	0.429	0.493	0.572	0.669	0.758	0.592	1.768
Egypt	2007	0.577	0.639	0.681	0.735	0.807	0.700	1.399
Ethiopia	2005	0.303	0.333	0.360	0.406	0.583	0.373	1.925
India	2005	0.547	0.581	0.644	0.699	0.756	0.615	1.380
Indonesia	2003	0.558	0.613	0.664	0.703	0.762	0.663	1.367
Kyrgyz Republic	1997	0.605	0.641	0.660	0.697	0.764	0.674	1.264
Nicaragua	2000	0.417	0.478	0.573	0.655	0.749	0.599	1.796
Nigeria	2003	0.291	0.390	0.454	0.541	0.632	0.466	2.170
Pakistan	2007	0.396	0.451	0.525	0.591	0.653	0.530	1.647
Peru	2005	0.533	0.610	0.654	0.748	0.805	0.686	1.508
Senegal	2005	0.326	0.367	0.432	0.505	0.617	0.462	1.891
Vietnam	2002	0.578	0.652	0.687	0.715	0.768	0.675	1.327
Zambia	2002	0.350	0.398	0.439	0.507	0.632	0.475	1.804

*Note:* HDI is based on the regression based approach for literacy and enrolment.

*Source:* Demographic and Health Surveys (DHS); calculations by the authors.

Table 5: HDI by urban and rural areas

Country	Year	By urban and rural areas			Ratio urban/rural
		Urban	Rural	Total	
Armenia	2005	0.781	0.753	0.770	0.964
Burkina Faso	2003	0.564	0.343	0.384	0.609
Bolivia	2003	0.659	0.465	0.592	0.706
Egypt	2007	0.753	0.659	0.700	0.875
Ethiopia	2005	0.562	0.348	0.373	0.618
India	2005	0.689	0.589	0.615	0.855
Indonesia	2003	0.705	0.626	0.663	0.888
Kyrgyz Republic	1997	0.721	0.651	0.674	0.903
Nicaragua	2000	0.675	0.487	0.599	0.721
Nigeria	2003	0.558	0.414	0.466	0.742
Pakistan	2007	0.612	0.487	0.530	0.796
Peru	2005	0.748	0.580	0.686	0.775
Senegal	2005	0.568	0.377	0.462	0.664
Vietnam	2002	0.742	0.663	0.675	0.894
Zambia	2002	0.584	0.413	0.475	0.707

*Note:* HDI is based on the regression based approach for literacy and enrolment.

*Source:* Demographic and Health Surveys (DHS); calculations by the authors.

Table 6: HDI by education of household head

Country	Year	Education of household head					Ratio higer/ no education
		No education	Primary	Secondary	Higher	Total	
Armenia	2005	0.683	0.754	0.761	0.818	0.770	1.198
Burkina Faso	2003	0.351	0.483	0.596	0.656	0.383	1.868
Bolivia	2003	0.444	0.540	0.648	0.733	0.591	1.650
Egypt	2007	0.591	0.702	0.733	0.792	0.700	1.340
Ethiopia	2005	0.343	0.391	0.483	0.600	0.373	1.748
India	2005	0.542	0.603	0.689	0.735	0.615	1.357
Indonesia	2003	0.540	0.636	0.717	0.754	0.663	1.395
Kyrgyz Republic	1997	0.580	0.614	0.671	0.717	0.674	1.235
Nicaragua	2000	0.475	0.590	0.701	0.748	0.599	1.574
Nigeria	2003	0.335	0.520	0.555	0.600	0.464	1.791
Pakistan	2007	0.456	0.552	0.595	0.638	0.529	1.400
Peru	2005	0.528	0.614	0.728	0.789	0.686	1.493
Senegal	2005	0.407	0.551	0.615	0.643	0.456	1.579
Vietnam	2002	0.543	0.630	0.705	0.766	0.675	1.410
Zambia	2002	0.344	0.443	0.538	0.629	0.475	1.830

*Note:* HDI is based on the regression based approach for literacy and enrolment.

*Source:* Demographic and Health Surveys (DHS); calculations by the authors.

Table 7: HDI by age of household head

Country	Year	Age of household head					Ratio oldest/ youngest
		20-29	30-39	40-59	60+	Total	
Armenia	2005	0.748	0.807	0.759	0.774	0.770	1.034
Burkina Faso	2003	0.356	0.381	0.391	0.392	0.384	1.098
Bolivia	2003	0.559	0.602	0.604	0.565	0.593	1.010
Egypt	2007	0.635	0.702	0.719	0.629	0.700	0.990
Ethiopia	2005	0.307	0.364	0.404	0.374	0.373	1.219
India	2005	0.552	0.629	0.620	0.615	0.615	1.115
Indonesia	2003	0.601	0.680	0.672	0.611	0.663	1.017
Kyrgyz Republic	1997	0.581	0.692	0.705	0.610	0.674	1.050
Nicaragua	2000	0.538	0.618	0.612	0.586	0.600	1.089
Nigeria	2003	0.432	0.472	0.485	0.464	0.466	1.073
Pakistan	2007	0.437	0.519	0.561	0.506	0.530	1.159
Peru	2005	0.611	0.694	0.706	0.661	0.686	1.082
Senegal	2005	0.452	0.475	0.500	0.459	0.462	1.015
Vietnam	2002	0.585	0.694	0.689	0.632	0.675	1.080
Zambia	2002	0.409	0.495	0.509	0.431	0.476	1.053

*Note:* HDI is based on the regression based approach for literacy and enrolment.

*Source:* Demographic and Health Surveys (DHS); calculations by the authors.

Table 8: HDI by sex of household head

Country	Year	Sex of household head			Ratio female/male
		Male	Female	Total	
Armenia	2005	0.775	0.761	0.770	0.982
Burkina Faso	2003	0.378	0.448	0.384	1.186
Bolivia	2003	0.589	0.609	0.592	1.034
Egypt	2007	0.701	0.682	0.700	0.973
Ethiopia	2005	0.364	0.415	0.373	1.140
India	2005	0.618	0.598	0.615	0.968
Indonesia	2003	0.666	0.628	0.663	0.944
Kyrgyz Republic	1997	0.670	0.689	0.674	1.028
Nicaragua	2000	0.588	0.626	0.599	1.064
Nigeria	2003	0.455	0.534	0.466	1.174
Pakistan	2007	0.527	0.553	0.530	1.048
Peru	2005	0.686	0.686	0.686	0.999
Senegal	2005	0.446	0.514	0.462	1.152
Vietnam	2002	0.674	0.680	0.675	1.009
Zambia	2002	0.477	0.470	0.475	0.985

*Note:* HDI is based on the regression based approach for literacy and enrolment.

*Source:* Demographic and Health Surveys (DHS); calculations by the authors.



Table 9: HDI by household size

Country	Year	Household size				Ratio large/small
		1-5	6-11	11+	Total	
Armenia	2005	0.770	0.771	0.734	0.770	0.953
Burkina Faso	2003	0.388	0.381	0.387	0.384	0.998
Bolivia	2003	0.617	0.577	0.555	0.592	0.900
Egypt	2007	0.701	0.704	0.622	0.700	0.888
Ethiopia	2005	0.352	0.380	0.433	0.373	1.231
India	2005	0.616	0.615	0.616	0.615	1.000
Indonesia	2003	0.655	0.671	0.669	0.663	1.023
Kyrgyz Republic	1997	0.689	0.668	0.629	0.674	0.913
Nicaragua	2000	0.626	0.592	0.516	0.599	0.824
Nigeria	2003	0.464	0.470	0.441	0.466	0.950
Pakistan	2007	0.505	0.536	0.527	0.530	1.042
Peru	2005	0.694	0.683	0.664	0.686	0.958
Senegal	2005	0.507	0.457	0.456	0.462	0.900
Vietnam	2002	0.686	0.667	0.619	0.675	0.902
Zambia	2002	0.438	0.487	0.540	0.475	1.232

*Note:* HDI is based on the regression based approach for literacy and enrolment.

*Source:* Demographic and Health Surveys (DHS); calculations by the authors.

Table 10: Inequality in the HDI by country

Country	Year	HDI 2	Gini index					Gini index (without log transformation)	GDP index capped	Gini index (PovcalNet) (%)
			Education index	Life index	GDP index	Income (uncapped)				
Armenia	2005	0.06	0.05	0.12	0.09	0.34	0.35	33.8	33.8	33.8
Burkina Faso	2003	0.21	0.44	0.28	0.19	0.40	0.44	39.6	39.6	39.6
Bolivia	2003	0.15	0.17	0.19	0.21	0.54	0.57	60.24	60.24	60.24
Egypt	2007	0.09	0.15	0.13	0.08	0.31	0.31	32.14	32.14	32.14
Ethiopia	2005	0.19	0.33	0.32	0.11	0.24	0.26	29.76	29.76	29.76
India	2005	0.11	0.24	0.13	0.09	0.30	0.31	33.32	33.32	33.32
Indonesia	2003	0.10	0.14	0.17	0.09	0.28	0.29	30.23	30.23	30.23
Kyrgyz Republic	1997	0.08	0.06	0.17	0.12	0.33	0.35	35.98	35.98	35.98
Nicaragua	2000	0.14	0.22	0.15	0.19	0.49	0.51	50.3	50.3	50.3
Nigeria	2003	0.21	0.28	0.31	0.22	0.43	0.47	42.93	42.93	42.93
Pakistan	2007	0.15	0.27	0.22	0.10	0.31	0.32	31.18	31.18	31.18
Peru	2005	0.11	0.12	0.17	0.15	0.49	0.50	53.01	53.01	53.01
Senegal	2005	0.19	0.36	0.23	0.15	0.39	0.41	39.19	39.19	39.19
Vietnam	2002	0.09	0.13	0.12	0.12	0.33	0.34	37.55	37.55	37.55
Zambia	2002	0.17	0.18	0.27	0.24	0.42	0.47	42.08	42.08	42.08

*Note:* HDI 1 is based only on literacy; HDI 2 is based on the regression based approach for literacy and enrolment. HDI 3 is based on the regression based approach for enrolment and on years of schooling per household aged +25.

*Source:* Demographic and Health Surveys (DHS); calculations by the authors.

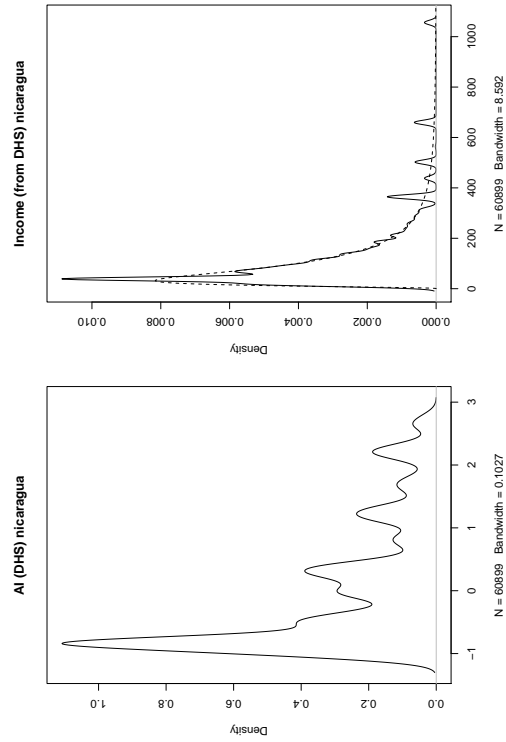
Table 11: Inequality decomposition of the HDI by subgroups

	By income quintiles			By urban and rural areas			By Education of household head		
	Total	Within group	Between group	Total	Within group	Between group	Total	Within group	Between group
Armenia (2005)	0.006	0.005	0.001	0.006	0.006	0.000	0.006	0.005	0.001
Burkina Faso	0.100	0.061	0.039	0.100	0.069	0.030	0.100	0.069	0.030
Bolivia	0.052	0.031	0.021	0.052	0.038	0.013	0.052	0.037	0.015
Egypt (2007)	0.025	0.016	0.009	0.025	0.022	0.003	0.025	0.014	0.011
Ethiopia (2005)	0.096	0.066	0.031	0.096	0.072	0.025	0.096	0.071	0.025
India (2005)	0.025	0.018	0.007	0.025	0.023	0.002	0.025	0.019	0.006
Indonesia (2003)	0.037	0.026	0.011	0.037	0.033	0.004	0.037	0.023	0.014
Kyrgyz Republic (1997)	0.010	0.007	0.002	0.010	0.009	0.001	0.010	0.009	0.001
Nicaragua (2000)	0.060	0.031	0.030	0.060	0.042	0.018	0.060	0.037	0.023
Nigeria (2003)	0.117	0.070	0.046	0.117	0.102	0.014	0.118	0.067	0.051
Pakistan (2007)	0.057	0.035	0.023	0.057	0.048	0.009	0.057	0.039	0.018
Peru (2005)	0.032	0.020	0.012	0.032	0.022	0.010	0.032	0.022	0.010
Senegal (2005)	0.084	0.044	0.039	0.084	0.053	0.030	0.084	0.057	0.026
Vietnam (2002)	0.024	0.019	0.006	0.024	0.023	0.001	0.024	0.016	0.008
Zambia (2002)	0.075	0.045	0.030	0.075	0.056	0.019	0.075	0.053	0.022

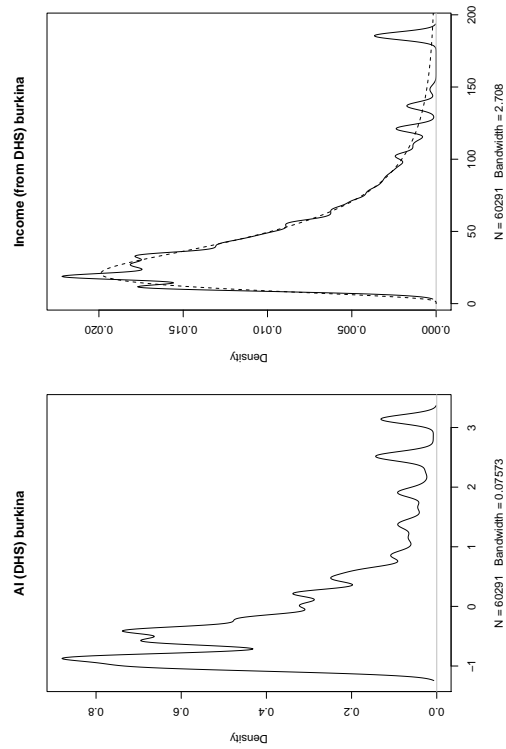
*Note:* HDI is based on the regression based approach for literacy and enrolment.

*Source:* Demographic and Health Surveys (DHS); calculations by the authors.

Figure 1: Distributions of the asst index and income  
(a)

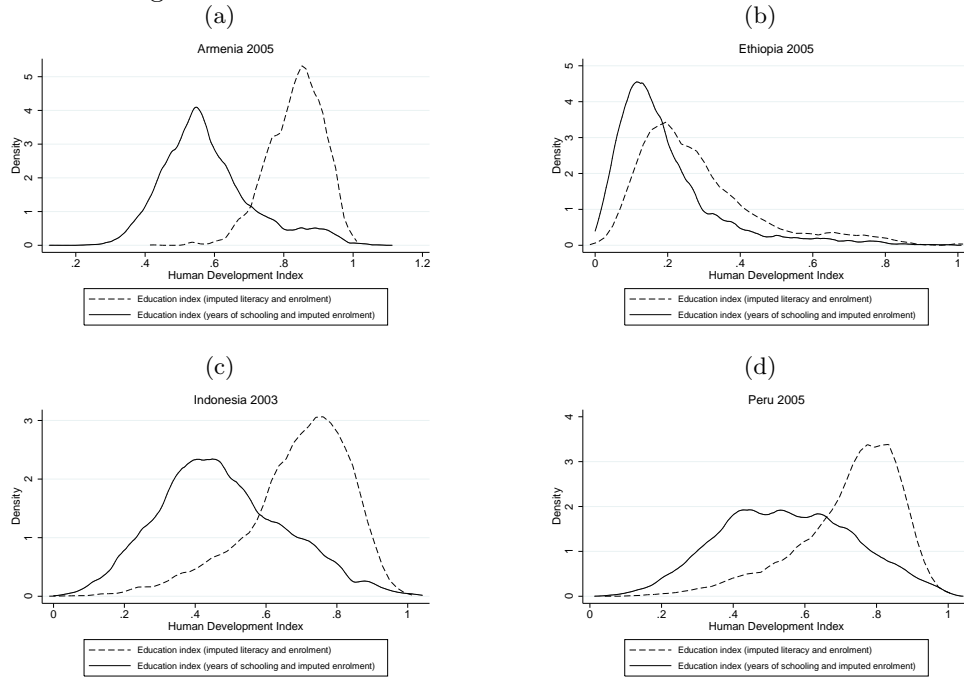


(b)



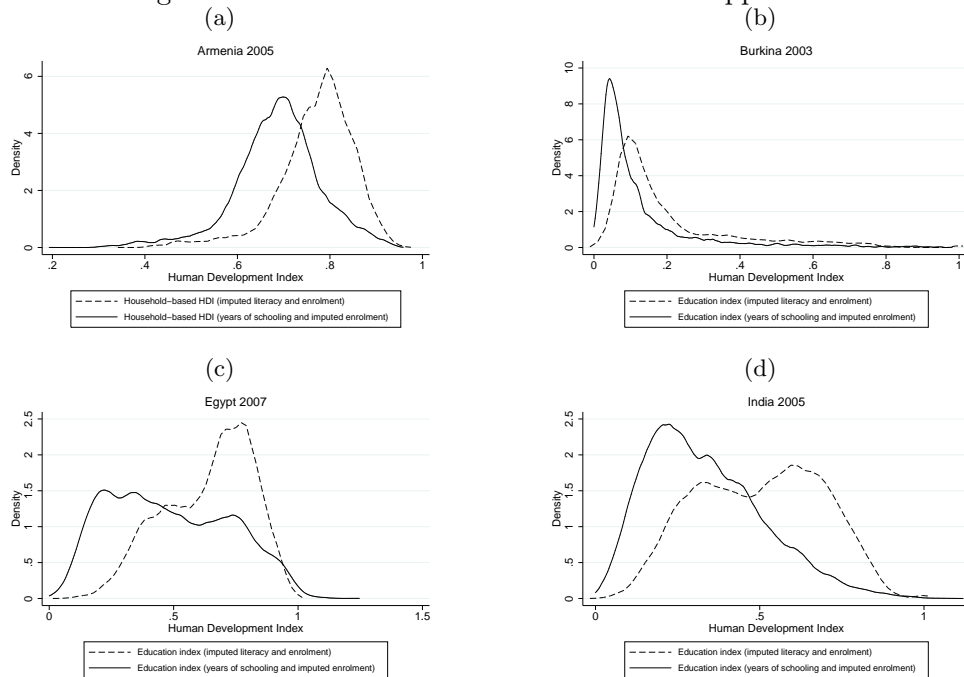
*Source:* Demographic and Health Surveys (DHS), PocvalNet; calculations by the authors.

Figure 2: Distributions of the alternative education indices



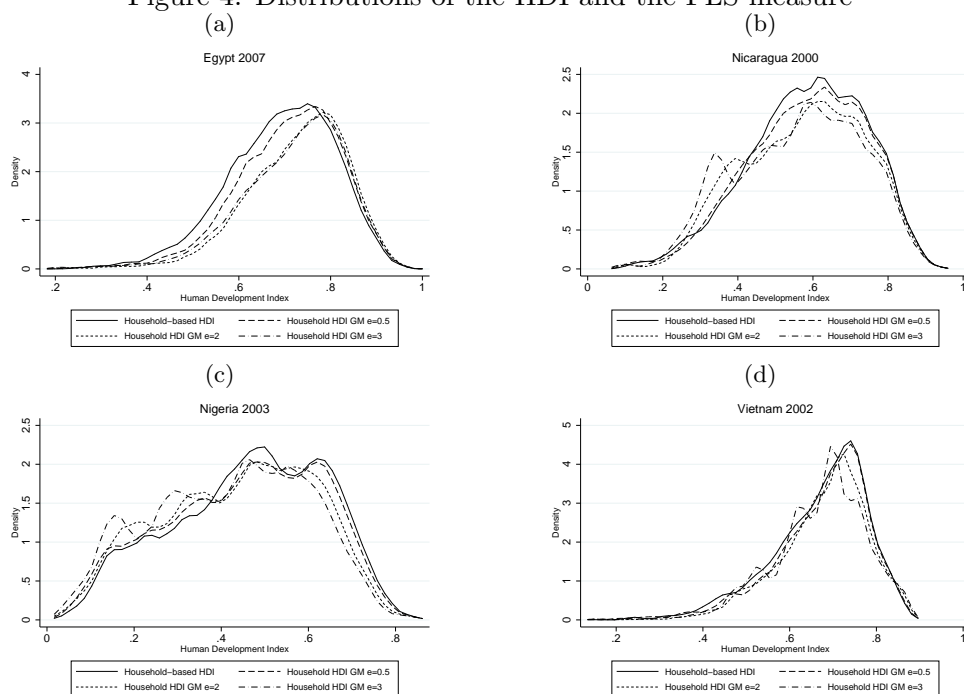
Source: Demographic and Health Surveys (DHS); calculations by the authors.

Figure 3: Distributions of the alternative HDI approaches



Source: Demographic and Health Surveys (DHS); calculations by the authors.

Figure 4: Distributions of the HDI and the FLS measure



Source: Demographic and Health Surveys (DHS); calculations by the authors.