Classification, Detection and Consequences of Data Error: Evidence from the Human Development Index

Hendrik Wolff
University of Washington

Howard Chong
Cornell University, hc757@cornell.edu

Maximilian Auffhammer
University of California - Berkeley

Follow this and additional works at: http://scholarship.sha.cornell.edu/articles

Part of the Real Estate Commons, and the Social Statistics Commons

Recommended Citation

This Article or Chapter is brought to you for free and open access by the School of Hotel Administration Collection at The Scholarly Commons. It has been accepted for inclusion in Articles and Chapters by an authorized administrator of The Scholarly Commons. For more information, please contact hlmdigital@cornell.edu.
Classification, Detection and Consequences of Data Error: Evidence from the Human Development Index

Abstract
We measure and examine data error in health, education and income statistics used to construct the Human Development Index. We identify three sources of data error which are due to data updating; formula revisions; and thresholds to classify a country's development status. We propose a simple statistical framework to calculate country specific measures of data uncertainty and investigate how data error biases rank assignments. We find that up to 34% of countries are misclassified and, by replicating prior studies, we show that key estimated parameters vary by up to 100% due to data error.

Keywords
data error, statistics, Human Development Index, HDI classification, data uncertainty, health, education, income

Disciplines
Real Estate | Social Statistics

Comments
Required Publisher Statement
Classification, Detection and Consequences of Data Error: Evidence from the Human Development Index

Hendrik Wolff, Howard Chong and Maximilian Auffhammer

We measure and examine data error in health, education and income statistics used to construct the Human Development Index. We identify three sources of data error which are due to data updating; formula revisions; and thresholds to classify a country’s development status. We propose a simple statistical framework to calculate country specific measures of data uncertainty and investigate how data error biases rank assignments. We find that up to 34% of countries are misclassified and, by replicating prior studies, we show that key estimated parameters vary by up to 100% due to data error.

Perhaps the greatest step forward that can be taken, even at short notice, is to insist that economic statistics be only published together with an estimate of their error.

- Oskar Morgenstern, 1970

This article studies the human development index (HDI), which has become one of the most widely used measures to communicate a country’s development status. Compared to the gross domestic product (GDP), the HDI is a broader measure of development, as it captures not only the level of income, but also incorporates measures of health and education (Srinivasan, 1994; Streeten, 1994; Anand and Sen, 2000). The United Nations Development Programme, which releases the HDI statistics, classifies each country into one of three categories: ‘low human development’ for HDI scores between 0.0 and 0.5, ‘medium human development’ for scores between 0.5 and 0.8 and ‘high human development’ for scores between 0.8 and 1.0.

Although these development categories were not originally designed to determine international relations, development aid, nor should they imply any other legal consequences, today these three mutually exclusive categories are used in politics, academia and the corporate world. In business relations, the categories have been used for international pricing purposes (Bate and Boateng, 2007).
Since 2001 the pharmaceutical company Merck sells drugs at different prices with up to 90% discounts for countries that are classified as ‘low’, and 75% reductions for ‘medium’ countries (Petersen and Rother, 2001). Second, the HDI has been widely used in debates among development researchers and policy makers (Sen, 2000) and is actively invoked to structure discussions in development-political debates of both governmental and non-governmental organizations (NGOs) (HDR, 1990–2006; Jahan, 2000). For allocation of development aid, it is known that the government of Ireland puts a particular focus on countries categorized as ‘low human development’ (O’Neill, 2005). International climate accord designs following the expiring Kyoto Agreement have included a proposal for linking countries’ abatement responsibilities according to their HDI (Hu, 2009). Thirdly, in economics, an extensive literature has studied the relationships between HDI rankings, economic growth, institutions, and other economic and social measures (Anand and Ravallion, 1993; Easterlin, 2000; Dasgupta, 2001). The conceptual underpinnings of the HDI can be found in the work by Amartya Sen (1977, 1984, 1985, 1987). For a recent mathematical ethical rationalization of the HDI, see Moreno-Ternero and Romer (2006). Oswald (1997), Blanchflower and Oswald (2005) and Leigh and Wolfers (2006) explore links between a happiness index and the HDI.¹

Despite extensive use of the HDI statistics, the drastic changes in the distribution of HDI scores for developing countries, as displayed in Figure 1, have gone unnoticed in the academic and policy literature. When the HDI was first published in 1990, the cross-country distribution appeared to be approximately uniformly distributed between zero (least developed) and one (most developed). Today, however, the distribution is twin-peaked with two sharp spikes around the values of 0.5 and 0.8, which are the cut-off values for categorizing countries of ‘low’, ‘medium’ and ‘high’ human development.

In this article, we investigate the role of data error on the published HDI and the consequences for its use in statistical analysis. We address these questions by exploiting: the originally published HDI time series; the sub-indicator variables used to construct the HDI; changes to the HDI formula; and documented data revisions. We identify three sources of data error: measurement error due to data revisions; data error due to formula updating; and misclassification due to inconsistent cut-off values. After isolating data revision error from error due to formula updates, we estimate country specific variances of the HDI scores. For example, the variance due to data revision for Bolivia represents the distribution of possible HDI values for Bolivia in a given year, which is solely created by updates to the

data series. We show that the HDI contains data error standard deviations ranging from 0.03 (US) to 0.11 (Niger), which is significant given the 0 to 1 scale. We find that the magnitude of the error variances is greater the lower the HDI rank, which is consistent with the quality of the statistical agencies improving with higher development. Likewise, country specific variances due to formula revisions are calculated. Mapping these cardinal noise measures onto the ordinal dimension, we find that 11, 21 and 34% of all countries can be interpreted as currently misclassified in the development bins due to the three sources of data error, respectively.

We also investigate the ordinal rank error. Each year when the new HDI statistics are published, much public attention focuses on the relative rank of a country to its rank in prior years and to the rank position of competing countries. For example, when Canada lost the top HDI number 1 position in 2001, The National Post (3rd of July 2001) wrote: ‘We’re not No. 1! Canada drops in UN rankings...Prime Minister Jean Chretien often refers to the report in public statements and speeches...’ Or, in 1998, when Pakistan (rank 138) bypassed India (rank 139), The Tribune (14th September 1998) noted: ‘Pak beat
To investigate the reliability of such statements, we calculate each country’s likelihood of deviation from its original published HDI rank. We find that on average the expected absolute deviation is nine rank positions. Furthermore, the average 95% confidence interval of our simulated HDI rank deviations ranges from -21 ranks to +20 ranks for the 2.5% percentile and the 97.5% percentile, respectively. These calculations show that statements based on ordinal comparisons are to be interpreted with great care.

Our results have direct implications for the academic literature. First, there is a vast economic literature that uses the same country level data that are included in the construction of the HDI, namely purchasing power parity adjusted income (Rogoff, 1996), life expectancy (Acemoglu and Johnson, 2007) and the educational measures of literacy rate and school enrolment statistics (Krueger and Lindahl, 2001). We investigate the inherent noise characteristics for each of these variables separately by estimating country specific variances for the underlying variables – GDP per capita, school enrolment, literacy rate and life expectancy. We find that the variables of health and education exhibit particularly large error variances in developing countries; in comparison income has a smaller error variance but among the three sub-indicators it reveals the largest updating bias. Second, the HDI has been used to analyze the evolution of the world’s distribution of well being, to explore issues of inequality, polarization, foreign direct investment, development aid and to test various convergence hypotheses in macroeconomics econometrically. By replicating some of these studies and carrying out sensitivity analyses, we find that key parameters, such as estimated Gini coefficients and speed of convergence parameters, vary by up to 100% in their values solely due to the measurement error.

Our article is related to the literature that discusses the challenges in accurately estimating national accounts and other aggregate statistics. Deaton and Heston (2008) provide an in depth analysis of the various factors that affect purchasing power parity (PPP). In their case, in order to eliminate differences in national price levels, GDP is combined with data by the International Comparison Programme (ICP) but the ICP’s methodologies are subject to various changes, (i.e. modifications of baskets, Laspeyres versus Paasche index, product quality adjustments). In discussing previous revisions of the methodologies, Deaton and Heston (2008) conclude that PPP data are ‘not always suitable for the purposes to which they are put’. Krueger and Lindahl (2001) study the relationship between economic growth and country level educational variables and discuss the direction of bias one might expect by using different variables. Other articles that characterize the noise in aggregate statistical data include

---

2 Pakistan ranked 119 and 138 and India ranked 118 and 139 in 1997 and 1998 respectively. For an extended discussion about these and similar rank statements see Morse (2003).
Barro and Lee (1993; 2001) and de la Fuente and Doménech (2006) for educational measures, Dowrick and Quiggin (1997) and Neary (2004) for income based measures and Anderson (1999) for life expectancy. We add to this literature by systematically isolating the different sources of error into data based errors, formula based errors and cut-off value based errors. To our knowledge, this is the first article to calculate country specific variance measures of the HDI, income, life expectancy, literacy rate, school enrolment, as well as to calculate indicators and probabilities of a country’s misclassification.

The remainder of the article is structured as follows. Section 1 describes the data. Section 2 outlines the framework and methods of measuring variances and misclassifications due to data revisions, formula changes and the threshold problem. Section 3 presents our results. Section 4 provides examples of how the HDI is used in various contexts and how errors can affects prior academic analysis. We conclude with policy recommendations in Section 5.

**Data**

The HDI is a composite indicator measuring a country’s level of development along three dimensions: health, education and income. These dimensions are expressed as unit-free and double-bounded sub-indicators $y_1, y_2, y_3$, each taking values between zero and one. The sub-indicators themselves are functions of data $x$ on primary and secondary school enrolment statistics, literacy rate, life expectancy and GDP per capita adjusted by PPP. Finally, the HDI is calculated as a simple average of the three ($k = 1, 2, 3$) sub-indicators, $\text{HDI} = \frac{1}{3} \sum_{i=1}^{3} y_i(x)$, which is then used for ordinal and cardinal comparisons. The HDI is published annually in the Human Development Reports (HDR) by the United Nations Development Programme (UNDP), which are available for the years 1990 to 2006 (HDR, 1990–2006).

**Original versus Revised Data**

In our analysis, we exploit the fact that the original historical data matrix $X_t$ used by the UNDP in year $t$ differs from the revised matrix $X^R_t$, which includes updates between $t$ and $s > t$. The original $X_t$ is available for the years $t = 1999–2006$, whereas the revised data $X^R_t$ are available: for all years of the analyses, $t = 1990–2006$ and $s = 2006$; and for the HDI in $t = 1975$, the revised $\text{HDI}^R_{1975}$ is available for $s = 3$.

---

3 The UNDP mainly draws the GDP data from the World Bank, the educational statistics are provided by UNESCO and life expectancy comes from the Population Division of the UN Department of Economic and Social Affairs. As countries do not consistently provide data using the same methodologies, these data sets are complemented by data from the Penn World Tables as well as by UNDP’s own estimates to impute missing values. See the technical appendices of the HDR (1990–2006) for details.
1999, 2000, ..., 2006. In this article, \( x_t^R \) refers to the variables for year \( t \) kindly provided to the authors as of fall of 2006 by the UNDP office, except stated otherwise. \( x_t \) refers to the data that we hand-copied\(^4\) from the \( t \)th year Human Development Report (HDR, 1990–2006).

**The HDI Formulas and Computation of Counterfactuals**

Since 1990, the UNDP has made three major updates to the formula used to construct the HDI. For each year \( t \) and country \( i \) the HDI formula is given by

\[
\text{HDI}_{it} = h_f(x_{it}).
\]

The formula \( h \) changed thrice as indexed by \( f \in \{1, 2, 3\} \), which corresponds to the time periods 1990, 1995–1998 and 1999–2006, respectively. The three formulas are explained in the HDR Technical Appendices (1990–99) of Jahan (2000) and in the Appendix of Wolff et al. (2010), the web-based version of this article.\(^5\) We construct three ‘counterfactuals’ denoted by \( h_A, h_{B_{ift}}, h_C \). Hence, for the entire time series we recalculate what the HDI would have been if the alternate formulas had been in place, using the most recent available historical data on the sub-indicators. In the analysis, we exploit exactly these differences between the ‘original’ HDI generated by the formula that was active at time \( t \) compared to the HDI generated by the other two formulas that were not active in that particular year \( t \).

**The Sample**

We construct a balanced panel from 1990 to 2006. A country is included in our panel if it meets the following two conditions: the country exists continuously between 1990 and 2006 (e.g. Croatia is dropped); and between the three revised sub-indicators and the countries’ HDI as provided by the UNDP, the total sum of missing data points is less or equal to five. Furthermore, in some of our analysis we distinguish between industrialized and non-industrialized countries whereby the industrialized countries are defined as in Table 1.1 of HDR (1991). We impute any missing data points by linear interpolation. In this way, we obtain a balanced panel 99 countries of which 76 are non-industrialized countries and 23 are industrialized countries.

---

\(^4\) The data were hand-copied separately by two of the authors. Only after verifying that the two hand-copied data sets are 100% identical, we proceeded with the analysis.

\(^5\) Wolff et al. (2010), the web-based version of this article, is available at http://faculty.washington.edu/hgwolff/EJOnlineWebVersionofHDI_Wolffetal2010.pdf.
Sources of Data Error and Methodology to Measure Data Uncertainty

This Section provides a detailed discussion of the three sources of data error: measurement error due to data revisions (D), data noise due to formula updating (F) and misclassification due to inconsistent cut-off values (C), which we abbreviate by D, F and C. We propose a simple statistical framework to analyze these sources of error, which allows us to calculate country specific variances and confidence intervals and to simulate country specific probabilities of misclassification.

Before discussing the details of each source of error below, it is useful to illustrate when the different types of errors (D, F and C) enter into the construction of the HDI. The columns of Table 1 show the overall structure of the data and the rows display when each error category contributes to the data uncertainty, depending on the level of data analyzed. The first column shows that with respect to the primary data variables $x$, the only source of error is due to data updating (D). For the sub-indicator functions $y$, two sources of errors are identified. First, with respect to D, $y$ is vulnerable because the data error of $x$ is directly translated into $y$ through the function $y(x)$. Additionally, the nonlinear functions $y(x)$ are subject to formula changes (F) over time. Similarly, the aggregate HDI measure is subject to D and F through $\text{HDI} = 1/3 \sum y_k(x)$. The HDI development categories are subject to error type C. Finally, the three types of error can be calculated for any function of HDI, $\text{��}(\text{HDI})$, e.g. Gini coefficients or regression parameters.

As we will make clear below, we calculate the three types of error independent of each other. Hence, it is not the case that error measure F will implicitly include some data error D or vice versa. Only in Section 2.4 and 3.2, we show how the different type of errors add up and discuss the correlation structures among them.

What are the distinctions between these sources of errors? While the first error D is well known to econometricians as ‘measurement error’, the changes to the data by F and C are due to subjective decisions by the data provider (here the UNDP). This subjective component changed over time and impacted the construction and relative importance of sub-variables of the HDI as well as the judgment on how to classify countries. Another distinction between D, F and C is that our first two types of errors, D and F, are cardinal in nature. This is in contrast to our third type of error, C, which is purely ordinal in nature in the sense that countries are either misclassified or not within the UN triplebin classification system.
To obtain the first measure of the randomness of the HDI data, we exploit the following exogenous changes to the data over time: the data $x_t$ (as used by the UNDP for the HDR at year $t$) are in general not the same data as the UNDP publishes in year $s$ for the same data year $t$. As revised statistics become available, the UNDP updates the original data matrix $x_t$ at year $s$, $s \geq t$, which we then denote $x^R_t$.

There are literally hundreds of reasons for data updates each year. The HDI draws their datasets from a multitude of domestic and international agencies (UNESCO, World Bank, Penn Tables). Often an agency may have data only for some subset of countries and some subset of years. The remaining data points are then filled by datasets from other agencies and occasionally are interpolated by neighboring years or countries. The dozens of footnotes in the yearly HDR reports point to the institutions that changed data year by year. The complexity of the problem may be best illustrated with a specific example: since 1999, the UNDP publishes historical HDI scores going back until the year 1975, HDI$_{1975}$. Figure 2 displays HDI$_{1975}$ scores as they are reported in each of the HDR reports from 1999 to 2006. In every year, between 1999 and 2006, substantial data revisions took place for the same 1975 HDI score. For example, while in 2000 Portugal was reported to have a historical HDI$_{1975}$ of 0.73 (that was below the HDI$_{1975}$ of Venezuela), by 2006 Portugal’s HDI$_{1975}$ increased to 0.79 and is now substantially above the 2006 reported HDI$_{1975}$ of Venezuela. On average across all countries the HDI updating bias for the year 1975 can be calculated as 0.003 with a standard deviation of the updating error of $\sigma_{1975} = 0.012$. Given that the data updates took place after a quarter of a century, we consider 0.012 to be large. Instead, in a world of good data quality $r_{1975}$ should be close to zero. This implies that whenever an analyst uses UNDP data, the same analysis run at a later date will result in different estimates due to a changed data matrix. Hence, when the HDI is released in year $t$, the value must be understood as an inexact value.

### Table 1

<table>
<thead>
<tr>
<th>Type of error</th>
<th>$x$</th>
<th>$y(x)$</th>
<th>HDI($y$)</th>
<th>$\theta$(HDI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data revisions D</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Formula updates F</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cut-off value C</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

**Notes.** For each column we indicate by the symbol (✓) which type of data error can affect the particular variable displayed in the column. $x$ refers to the raw variables, $y(x)$ to the sub-indicators which are functions of $x$. HDI is a function of the $y$ and $\theta$(HDI) refers to any parameter of interest (i.e. Gini coefficient) that is calculated as a function of one or multiple HDI values.

**First Source of Data Error: Measurement Error**

To obtain the first measure of the randomness of the HDI data, we exploit the following exogenous changes to the data over time: the data $x_t$ (as used by the UNDP for the HDR at year $t$) are in general not the same data as the UNDP publishes in year $s$ for the same data year $t$. As revised statistics become available, the UNDP updates the original data matrix $x_t$ at year $s$, $s \geq t$, which we then denote $x^R_t$.

There are literally hundreds of reasons for data updates each year. The HDI draws their datasets from a multitude of domestic and international agencies (UNESCO, World Bank, Penn Tables). Often an agency may have data only for some subset of countries and some subset of years. The remaining data points are then filled by datasets from other agencies and occasionally are interpolated by neighboring years or countries. The dozens of footnotes in the yearly HDR reports point to the institutions that changed data year by year. The complexity of the problem may be best illustrated with a specific example: since 1999, the UNDP publishes historical HDI scores going back until the year 1975, HDI$_{1975}$. Figure 2 displays HDI$_{1975}$ scores as they are reported in each of the HDR reports from 1999 to 2006. In every year, between 1999 and 2006, substantial data revisions took place for the same 1975 HDI score. For example, while in 2000 Portugal was reported to have a historical HDI$_{1975}$ of 0.73 (that was below the HDI$_{1975}$ of Venezuela), by 2006 Portugal’s HDI$_{1975}$ increased to 0.79 and is now substantially above the 2006 reported HDI$_{1975}$ of Venezuela. On average across all countries the HDI updating bias for the year 1975 can be calculated as 0.003 with a standard deviation of the updating error of $\sigma_{1975} = 0.012$. Given that the data updates took place after a quarter of a century, we consider 0.012 to be large. Instead, in a world of good data quality $r_{1975}$ should be close to zero. This implies that whenever an analyst uses UNDP data, the same analysis run at a later date will result in different estimates due to a changed data matrix. Hence, when the HDI is released in year $t$, the value must be understood as an inexact value.
subject to future data revisions. This problem is what we refer to as measurement error from data updating.

To parameterize this measurement error, assume that the relationship between the observed HDI score of country $i$ and the true (but unknown) sub-indicators, denoted by $\bar{x}^{*}$, can be expressed as

$$\text{HDI}_i = 1/3 \sum_k (\bar{y}_{ik}^* + \epsilon_{ik})$$

where $\bar{x}^{*}$ is orthogonal to $\bar{x}^k$ and is distributed with mean $m_{ik}$ and country specific variance $\sigma^2_{ik}$. The relationship between the observed HDI score of country $i$ and the true HDI* consequently is $H \bar{x}^{*} + e$ with $e_{it}$ being the composite error term distributed with mean $1/3 \sum_i m_{ik}$ and country specific variance $\sigma^2_i$ that is determined by the covariance structure of the measurement error of the sub-indicators in country $i$, $\text{cov}(\bar{x}^k)$.

Exploiting the original $x_t$ and revised $x_t^R$ we now are in the position to calculate country specific variances of the measurement error due to data updating (D) given by

$$\sigma_{D,t}^2 = \sum_t \left[ h_t(x_t) - h_t(x_t^R) \right]^2 / (T^* - 1)$$

for all $t \in T$.

The variance of the data-updating measurement error is based on the difference between the original HDI as published in the HDR at year $t$, $h_t(x_t)$, and the reconstructed counterfactual HDI for year $t$ using

![Fig. 2. HDI of 1975 of Portugal and Venezuela as Reported in the Years 1999 to 2006](image-url)
revised data \( x_{t}^{R} \) available to us today. To obtain a consistent estimate of the variance, we assume that 
\( h_{t}(x_{t}^{R}) \) represents our currently best available estimate of \( H_{t}^{R} \) and discuss in our result the implications of this assumption.

Importantly, note that we calculate \( \sigma^{2}_{D} \) independently from error type F. Specifically, we disentangle D from F by constructing each pair of data \([h_{t}(x_{it}^{R}), h_{t}(x_{it}^{R} R)]_{t} \forall t \in T\) to be conditional on the same HDI formula, namely the formula that was active at time \( t \). (Instead, if one were using the pairs of data \([ht(x_{it}), HDIR]\) as reported in the yearly UNDP reports, one would have erroneously incorporated error-type F into error type D).

**Second Source: Changes in HDI Formula**

Since its release in 1990, the HDI was often criticized with respect to its analytical framework and methodology (Desai, 1991; Kelley, 1991; McGillivray, 1991; Aturupane et al., 1994; Noorbakhsh, 1998). The UNDP responded to this challenge by working with Nobel laureate Amartya Sen, Sudhir Anand, Paul Streeter and others to intellectually lead an effort to update the methodology and value judgments. As a result UNDP has made three major updates to the formula used to construct the HDI which are further discussed in Anand and Sen (1994, 1997, 2000), Jahan (2000), the Technical Appendices of the HDRs (1990–2006) and summarized in the Appendix of the web-based version of this article (Wolff et al., 2010). These three changes are clearly visible in the empirical distribution of the HDI displayed in Figure 3. In particular, different distributional characteristics occur for the sub-periods A (1990), B (1995–98) and C (1999–2006) that correspond to the three formula regimes \( h_{A}(x_{it}), h_{B}(x_{it}), \) and \( h_{C}(x_{it}) \), respectively.

We exploit this variation of the HDI scores across the counterfactual formulas to calculate country specific variances due to the formula (F) updates that is

\[
\sigma^{2}_{F,j} = \frac{\sum g \left[ h_{g}(x_{it}^{R}) - h_{C}(x_{it}^{R}) \right]^{2}}{(2T^* - 1)} \forall t \in T
\]

where \( g \) is the index to sum over the formula indices A, and B. The variance \( \sigma^{2}_{F,j} \) is based on the country specific differences of the HDI generated by the most recent and improved formula \( h_{C} \) compared to the HDI counterfactuals generated by the other two formulas \( h_{A} \) and \( h_{B} \). We do acknowledge that the formula revisions were undertaken to improve the HDI statistics and hence one interpretation of \( \sigma^{2}_{F,j} \) is to understand it as a measure of historical noise due to the formula updates. Alternatively, the country
specific measures $\sigma^2_{F,t}$ can be interpreted as a present measure of noise, if the UNDP will similarly continue to change the formula in the future and the scores today would have to be understood as subject to those future formula revisions.

Note that we again isolate the error type F from the former error type D. Hence, it is not the case that error-type F incorporates error-type D, and/or vice versa.6

Third Source of Misclassification: Arbitrary Cutoff Values

In comparison to the cardinal measures of noise due to D and F, our third measure of error, C, is entirely ordinal. It is an error of misclassification due to the arbitrariness of the two cut-off values used to categorize countries into ‘low’, ‘medium’ and ‘high’ development countries. Despite the fact that changes made to the HDI formula did have considerable impacts on the empirical HDI distributions as displayed in Figure 3, the UNDP has decided to use the same cut-off values (0.5 and 0.8) since 1990. Since the original cutoff values are supposed to distinguish three qualities of human development, with each formula change the UNDP could instead have adjusted the cut-off values in such a way that the new adjusted thresholds again reflect these same value judgments for the levels of quality. One possible

---

6 We achieve the independence because the function $\sigma_F$ is defined conditional on the revised data $x^R_t$. Hence, all terms on the right hand side of $\sigma_F$ are ‘counterfactual’ measures, what the HDI would have been if the revised data $x^R_t$ had been already known in prior years under the different formula assumptions.
procedure\textsuperscript{7} to obtain revised threshold values – that are consistent with the initial 1990 value judgment of classifying quality and consistent with the entire history of formula changes – is as follows. In 1990, Morocco and Egypt were the two countries closest to the original cut off value of 0.5 (with HDI scores of 0.49 and 0.50, respectively). On the counterfactual distribution of formula $h_c$ applied to 1990, these two countries take on the values 0.54 and 0.56. Taking the mean (0.55) provides the revised threshold for separating between the low and medium human development groups. Similarly, we proceed with the cut off value 0.8 and obtain the revised value 0.70.

**Overall Error Variance**

So far, we have treated the two sources of errors D and F independently of each other. The user of the HDI statistics may, however, be also interested in having a sense of the ‘overall’ error within the HDI database.\textsuperscript{8} To this end, we calculate the country specific overall cardinal error variance statistics as

$$\sigma^2(\text{overall})_i = \sigma^2_D + \sigma^2_F + 2\text{cov}(e_D, e_F)$$

which takes into account the covariance structure of the individual error contributions,

$$\text{cov}(e_D, e_F) = \frac{\sum (e_{Di} - m_D)(e_{Fi} - m_F)}{T^*}$$

whereby $D_{\text{t}} = H_{\text{t}}(\text{original})^R - H_{\text{t}}(\text{updated})$ and $F_{\text{t}} = \sum Q^R h_f (\text{original})^F - h_c (\text{updated})/2$. We can thus analyze how much each source of error (1) and (2) contributes to the overall level of error in the HDI database.

**Simulation 1: The Expected Number of Misclassified Countries**

For the cardinal sources of data error, for each country we can calculate the probability of being misclassified. Given the parameterization of the measurement error as $H_{\text{t}}(\text{original})^* = H_{\text{t}}(\text{updated}) + e$ and assuming $\text{dist}(\text{original})^* \sim \text{dist}(0, \sigma^2_t)$, normally distributed with mean zero and variance $\sigma^2_t$ (as calculated by $\sigma^2_t$, $\sigma^2$, and $\sigma^2(\text{overall})_i$) we analytically calculate for each country the probability of being misclassified $\text{dist}(\text{original})_i$ as

\textsuperscript{7} Our procedure to choose the revised bin cutoffs is based upon the objective to maintain constant the initial (1990) value judgment by the UNDP, in the sense that the thresholds separate low from medium and medium
from high developed countries. One referee suggested selecting those cutoff values which maximize the objective function to maintain the development category of as many countries as possible. This would lead to the revised thresholds values of 0.62 and 0.76.

8 We thank the editor for providing the idea to aggregate errors.
where \( p() \) is the probability density function of the estimated *HDI* distributions. Hence, for countries reported to be ‘low development’, we calculate the probability of being classified as a medium or a high development country; similarly, we proceed for the ‘medium’ and ‘high’ development countries. Finally, adding these integrals over all countries provides the expected number of misclassified countries.

**Simulation 2: The Expected Number of Deviation in HDI Ranks**

In addition to sorting countries into the three broad HDI categories of ‘low’, ‘medium’ and ‘high’, the UNDP statistics are used to produce league rankings of countries. We calculate the expected number of absolute deviations in rank by simulating \( (n = 1, \ldots, 10,000) \) the 2006 HDI ranking. The simulated rankings are produced by calculating for every country \( i \) the simulated HDI as \( \text{SimHDI}_{i,2006} = \text{HDI}_{i,2006} + \eta_i \) with \( \eta_i \) distributed as mean zero and variance \( r^2(\text{overall})_i \). Finally, after each nth simulation country i’s simulated rank is recorded relative to its actual observed rank in 2006.

**Results**

**Results with Respect to the Cardinal Errors of Data Updating and Formula Changes**

If one followed Oskar Morgenstern’s (1970) advice given in the Introduction, an alternative way for UNDP to report HDI scores would be to report country specific noise measures. To do so, we display country specific standard errors in Table 2. With respect to the standard errors due to the measurement error of data updating (column 8), we find that \( \sigma_{D,i} \) ranges between a minimum value of 0.004 (US) and a maximum value of 0.069 (Syria), with an average value across all countries of 0.026. Given that the HDI is an average over three sub-indicators, whereby positive and negative deviations in the sub-indicators cancel out, and given that the HDI is scaled from 0 to 1, these standard deviations are large and significant. Figure 4 displays the relationship between the country specific measurement error due to

---

\[ \text{where } p() \text{ is the probability density function of the estimated } H \text{ distributions. Hence, for countries reported to be ‘low development’, we calculate the probability of being classified as a medium or a high development country; similarly, we proceed for the ‘medium’ and ‘high’ development countries. Finally, adding these integrals over all countries provides the expected number of misclassified countries.} \]

**Simulation 2: The Expected Number of Deviation in HDI Ranks**

In addition to sorting countries into the three broad HDI categories of ‘low’, ‘medium’ and ‘high’, the UNDP statistics are used to produce league rankings of countries. We calculate the expected number of absolute deviations in rank by simulating \( (n = 1, \ldots, 10,000) \) the 2006 HDI ranking. The simulated rankings are produced by calculating for every country \( i \) the simulated HDI as \( \text{SimHDI}_{i,2006} = \text{HDI}_{i,2006} + \eta_i \) with \( \eta_i \) distributed as mean zero and variance \( r^2(\text{overall})_i \). Finally, after each nth simulation country i’s simulated rank is recorded relative to its actual observed rank in 2006.

**Results**

**Results with Respect to the Cardinal Errors of Data Updating and Formula Changes**

If one followed Oskar Morgenstern’s (1970) advice given in the Introduction, an alternative way for UNDP to report HDI scores would be to report country specific noise measures. To do so, we display country specific standard errors in Table 2. With respect to the standard errors due to the measurement error of data updating (column 8), we find that \( \sigma_{D,i} \) ranges between a minimum value of 0.004 (US) and a maximum value of 0.069 (Syria), with an average value across all countries of 0.026. Given that the HDI is an average over three sub-indicators, whereby positive and negative deviations in the sub-indicators cancel out, and given that the HDI is scaled from 0 to 1, these standard deviations are large and significant. Figure 4 displays the relationship between the country specific measurement error due to

---

\[ \text{The correlation between the three sub-indicator error terms } \epsilon_{ik}, k \in \{1,2,3\} \text{ is close to zero and can be viewed as distributed approximately independently. Hence, the average standard deviation of the subindicator errors } \sigma_{D,i} \text{ must be larger in magnitude, compared to the standard deviation of the HDI, } \sigma_{D,i}. \text{ Section 3.4 confirms this by analyzing the compound error term.} \]
the data revisions, \( \sigma_{D,i} \), and the countries’ HDI score (as of 2006). We note that more developed countries have smaller updating variances. Similarly, column (3) displays the country specific data measurement errors due to formula updates \( \sigma_{F,i} \), whose ranges on average are even higher compared to \( \sigma_{D,i} \). We find the estimated \( \sigma_{D,i} \) range between a minimum value of 0.034 and a maximum value of 0.127 with a world average standard deviation of 0.072.

As the HDI is primarily used as an ordinal measure, we now turn to the impact of these cardinal measures on the ordinal dimension. Figure 5 displays the case of the ‘average’ non-industrial country with HDI = 0.65 using the average standard deviation over all non-industrialized countries due to data revisions, \( \sigma_{D}=0.03 \) and due to formula updates \( \sigma_{F}=0.08 \). Figure 5 shows that substantial probability mass is spread over all three development categories. In Table 2, the category specific probabilities are displayed for all countries in columns (4)–(6), and (9)–(11) for the formula based error and data upgrading errors, respectively. For example, as of 2006, South Africa, Mongolia, Syria, India, Honduras, Bolivia have non-zero probabilities of belonging to all three categories simultaneously. Even a high human development country, such as Costa Rica with HDI of 0.84, can still be a ‘low’ with 0.3% probability and yet be ‘medium’ to 37%. Finally, columns (7) and (12) display the total probability of a particular country being misclassified by using formula (4). The sum over these column probabilities show that currently, in expectation, 10.4 countries are misclassified due to data updating measurement error and 20.7 countries are misclassified due to formula updates; these numbers translate into, 11 and 21% of all countries being misclassified.

<table>
<thead>
<tr>
<th>Country i</th>
<th>2006 reported Human development status</th>
<th>2006 HDI</th>
<th>Measures based on formula updates (F)</th>
<th>Measures based on measurement error due to data revisions (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Niger</td>
<td>Low</td>
<td>0.31</td>
<td>0.13</td>
<td>0.05</td>
</tr>
<tr>
<td>Malawi</td>
<td>Low</td>
<td>0.33</td>
<td>0.13</td>
<td>0.05</td>
</tr>
<tr>
<td>Togo</td>
<td>Low</td>
<td>0.33</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>Uganda</td>
<td>Medium</td>
<td>0.33</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>South Africa</td>
<td>Medium</td>
<td>0.33</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>Brazil</td>
<td>High</td>
<td>0.33</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>Mauritius</td>
<td>High</td>
<td>0.33</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>High</td>
<td>0.33</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>Norway</td>
<td>High</td>
<td>0.33</td>
<td>0.12</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Table 2: Country i Specific Standard Deviations and Probabilities of Belonging to Development Category j

<table>
<thead>
<tr>
<th>Country i</th>
<th>2006 reported Human development status</th>
<th>2006 HDI</th>
<th>Measures based on formula updates (F)</th>
<th>Measures based on measurement error due to data revisions (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Niger</td>
<td>Low</td>
<td>0.31</td>
<td>0.13</td>
<td>0.05</td>
</tr>
<tr>
<td>Malawi</td>
<td>Low</td>
<td>0.33</td>
<td>0.13</td>
<td>0.05</td>
</tr>
<tr>
<td>Togo</td>
<td>Low</td>
<td>0.33</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>Uganda</td>
<td>Medium</td>
<td>0.33</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>South Africa</td>
<td>Medium</td>
<td>0.33</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>Brazil</td>
<td>High</td>
<td>0.33</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>Mauritius</td>
<td>High</td>
<td>0.33</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>High</td>
<td>0.33</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>Norway</td>
<td>High</td>
<td>0.33</td>
<td>0.12</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Notes: This Table is considerably shortened to nine countries only. The full set of countries is provided in Table 2 of the web-based version of this article (Wolff et al., 2010).

As the HDI is primarily used as an ordinal measure, we now turn to the impact of these cardinal measures on the ordinal dimension. Figure 5 displays the case of the ‘average’ non-industrial country with HDI = 0.65 using the average standard deviation over all non-industrialized countries due to data revisions, \( \sigma_{D}=0.03 \) and due to formula updates \( \sigma_{F}=0.08 \). Figure 5 shows that substantial probability mass is spread over all three development categories. In Table 2, the category specific probabilities are displayed for all countries in columns (4)–(6), and (9)–(11) for the formula based error and data upgrading errors, respectively. For example, as of 2006, South Africa, Mongolia, Syria, India, Honduras, Bolivia have non-zero probabilities of belonging to all three categories simultaneously. Even a high human development country, such as Costa Rica with HDI of 0.84, can still be a ‘low’ with 0.3% probability and yet be ‘medium’ to 37%. Finally, columns (7) and (12) display the total probability of a particular country being misclassified by using formula (4). The sum over these column probabilities show that currently, in expectation, 10.4 countries are misclassified due to data updating measurement error and 20.7 countries are misclassified due to formula updates; these numbers translate into, 11 and 21% of all countries being misclassified.
We interpret the misclassification of 11% due to data updating as conservative because $\sigma^2_{D,t}$ is just based on ‘short-term’ differences between $x_t$ and $x^R_t$, based on the years from 1990 to 2006. There is also ‘long-term’ data-updating error. Taking that into account may increase $\mathbf{t}_D$ as $\mathbf{t}_R$ increases with $s$. While we cannot capture this long-term effect by formula (1) (due to the lack of published original data prior to the HDR of 1990), we illustrated the magnitude of such ‘long-term’ drift in Figure 2.

Fig. 4. Relationship Between Countries’ Development Status and the Standard Deviations Due to Measurement Error Generated by Data Updates
Notes. Linear trendline based on sample of 99 countries, $R^2 = 0.184$.

Fig. 5. Representation of Data Error of a Country with HDI = 0.65
**Overall Cardinal Error and Rank Simulations**

The typical user of the HDI statistics may not be concerned about the individual error statistics $\hat{\sigma}_D$ and $\hat{\sigma}_F$ if they are calculated independently of each other but the researcher may be more interested in obtaining a sense of the overall error in the data. For this purpose we calculate country specific overall cardinal error statistics $\sigma^2(\text{overall})$ and find that the world average of these measures $\sigma^2(\text{overall}) = \frac{\sum \sigma^2(\text{overall})}{N}$ equals to 0.007, compared to $\hat{\sigma}_D = 0.001$ and $\hat{\sigma}_F = 0.006$. Furthermore, we find that all country specific covariance terms $\text{cov}(e_D, e_F)$ are relatively small (all correlation coefficients are smaller than 0.06 in absolute value) which implies that the updating error is not linearly correlated with the formula error. This implies that 86% of the total HDI variance is contributed by the formula error and 14% by the measurement error due to data updating.\textsuperscript{10} By using the same methodology as in Section 2.5, we calculate the ‘overall’ expected number of countries misclassified as 22.9. The country specific overall variance statistics are given in column (3) of Table 3.

Moreover, column (1) of Table 3 displays the country specific expected absolute value of rank displacements based upon the rank of the country’s HDI in 2006. Worldwide, the average country is displaced by about nine ranks. This average absolute displacement obscures the direction of rank displacement and the uncertainty over rank displacements. To this end, Figure 6 displays the average rank displacement over 10,000 simulations as a function of the countries’ 2006 HDI score along with the 95% confidence intervals. The confidence intervals are large, leading to an average deviation of -21 ranks and +20 ranks for the 2.5% percentile and the 97.5% percentile, respectively. Figure 6 also shows that countries with a low initial 2006 rank (low HDI score) do on average better in the simulated rank statistics and countries with an initial high HDI in 2006 are more likely to lose ranks in the simulations.

\textsuperscript{10} We calculate the percentage contribution of the $j$th cardinal source of error to the overall error as $\sigma^2_j / \sigma^2(\text{overall})$. This calculation is hence net of the covariance of the two error sources. The covariance terms can essentially be neglected due to the fact these are small in magnitude.
Results with Respect to the Cutoff Value Problem

Our third measure of misclassification is due to the non-adjustment problem of the cutoff values 0.5 and 0.8 that the UN uses to classify countries as low, medium and high human developed countries.

If the UNDP had adjusted the cut-off values in a manner consistent with the 1990 classification, since 1999 (the year of the last formula update), the thresholds should be at the values 0.55 and 0.70, as opposed to 0.5 and 0.8. This lack of adjustment of the cutoff values results in 34% of the countries being
misclassified today.\textsuperscript{11} Among all developing countries the percentage of misclassification is even higher: 45%. With such a high percentage statements such as ‘over the last decade x\% of African countries successfully moved from the “low” to the “medium” human development category’ – as expressed in numerous policy papers and news reports (United Nations, 1996; People’s Daily, 2001; Daily Times, 2005) – become useless at best, if not blatantly misleading.

**Measurement Error with Respect to the Underlying Variables of the HDI**

Thus far, we analyzed the data error of the HDI. As the same variables used to construct the HDI serve as key data in many academic studies as well as inputs to many other international comparative statistics, it is worthwhile analyzing the sub-indicators $y$ pertaining to health, education and income in more detail.

The first four columns of Table 4 display summary statistics of the overall HDI updating error, $e$, and the vector of sub-indicator updating errors, $\epsilon$, for our sample of 76 non-industrialized countries. In general, the standard deviations of the health and education indexes are larger than the standard deviations of the income statistics. It is interesting to note, however, that the main driver for the HDI upward bias stems from the change to the income index ($m_{\text{income}} = 0.01$).\textsuperscript{12} Instead, the errors on the health and the education indices show distributions that are centered around zero. Note, that the min / max columns in Table 4 reveal some enormous changes; the income index changed by 15% (Sudan and Chad) and the education index even by 25% (Mongolia) on the total scale from 0 to 1.

One may ask whether the three sub-indicator updating errors are correlated. An analysis of the year-by-year correlation matrices of the errors does not show any systematic co-movement, as the correlation coefficients are close to zero in all years. This suggests that the statistical adjustments on the three dimensions are independent of each other and indicates that the respective national statistical offices responsible for health, education, and income statistics have no systematic contemporaneous responses. Furthermore, statistical independence of the three sub-indicator error variables $\epsilon_{k}, k$ implies that their errors must be on average larger than the variance of the HDI error $e$, which is confirmed by Table 4. Hence, while the three sub-indicator errors offset each other with respect to the HDI,\textsuperscript{13} when working with the variables of education, income and health, one faces even larger data error.

\textsuperscript{11} The percentage of countries misclassified is calculated as the number of countries that have HDI scores in the ranges $[0.5, 0.55)$ and $[0.70, 0.8)$ divided by the total number of countries in our sample (99).

\textsuperscript{12} Statistically, this upward bias with a standard deviation of 0.02 is not significantly different from zero.

\textsuperscript{13} Under the assumption of independence, the standard deviation for the composite HDI error, $e$, is given by $\text{std}(e) = \sqrt{\sum s_k^2/9}$, which, after replacing $s_k$ by, equals to $\text{std}(e) = 0.014$. The estimated standard deviation of the
To analyze the drivers of the HDI data error in more detail, we calculate country specific noise measures due to data revisions with respect to the underlying variables, x. Table 5 reports country specific standard errors calculated as the country specific standard deviation \( \sigma(x_n) \) (computed analogously to (2) by exploiting the 2006 data revision of \( x_{nit} \) for \( t = 1999–2005 \)). In order to obtain a sense of the relative magnitude of the errors in each variable, we divide the standard deviations by the level corresponding variable in the year 2006, \( x_{n2006} \) and display the resulting relative standard errors in Figure 7. Adult literacy rate, GDP and the gross enrollment ratio contribute most to the updating error of the HDI. In contrast, life expectancy is revised much less. As is clearly recognizable in Figure 7, we find that the more highly developed the country the smaller its measurement error due to data updating.

Table 4

<table>
<thead>
<tr>
<th>Indicators</th>
<th>Non-industrialised Countries</th>
<th>Industrialised Countries</th>
<th>Industrial versus Non-industrialised Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
<td>Min</td>
</tr>
<tr>
<td>HDI</td>
<td>0.01</td>
<td>0.02</td>
<td>0.06</td>
</tr>
<tr>
<td>Health</td>
<td>0.00</td>
<td>0.03</td>
<td>0.14</td>
</tr>
<tr>
<td>Education</td>
<td>0.00</td>
<td>0.03</td>
<td>0.11</td>
</tr>
<tr>
<td>Income</td>
<td>0.01</td>
<td>0.02</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Notes. * States that estimate of ‘Differences in means’ is statistically significant at 1% level, tested by regressing the vector of updating error on a constant and an indicator variable that takes the value one if the country is industrialised and zero otherwise using robust standard errors. † Indicates that the estimate of the same regression is different from zero only at the 15% significance level.

12 Statistically, this upward bias with a standard deviation of 0.02 is not significantly different from zero.

Discussion of the Results

Given that the HDI is subject to a considerable amount of measurement error, the use of the HDI and its triple bin classification system can lead to serious interpretability problems. We now investigate the consequences of these three sources of errors by replicating prior studies and uses of the HDI, with each of the analysis being uniquely linked to our three sources of errors.

HDI measurement error by formula (1) applied to period C is 0.015, hence very close to std(e), confirming this theoretical result of independence.
The HDI as a Definitional Measure

While there does not exist a standardized definition of the term ‘developing country’, the definition is often linked to the HDI, as being a country with low to moderate development status. In fact, scientific studies have often been explicitly using the HDI system to identify a set of developing countries (Noorbakhsh, 2006; Varenne, 2007; Lauber and Roessler, 2007; Alvan, 2009). Leading online dictionaries do refer to the HDI in order to define the term ‘developing country’ (Wikipedia, 2008; Babylon, 2009; SearchWiki, 2009). Here, it is common to differentiate development status by using three different colors. In Figure 8, we recreate such a map by displaying the HDI scores for 2006. To demonstrate the impact of misclassification in our sample, we reclassify the countries using the updated thresholds of 0.55 and 0.70 as discussed in Section 2.3. The visual impact of this reclassification is striking, especially in South America, Southeast Asia and Africa. This misclassification is particularly problematic, if organizations/institutions use these categories to design particular policies or rules.
Table 5
Upgrading Error Statistics of Variables Underlying the HDI

<table>
<thead>
<tr>
<th>Country</th>
<th>HDI 2006</th>
<th>GDP per capita 2006 (PPP US$)</th>
<th>σ(GDP)_2006</th>
<th>σ(GER)_2006 (%)</th>
<th>σ(ALR)_2006 (%)</th>
<th>Life expectancy (LE) at birth 2006 (years)</th>
<th>σ(LE)_2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Niger</td>
<td>0.511</td>
<td>779.1</td>
<td>67.7</td>
<td>21.5</td>
<td>0.6</td>
<td>44.6</td>
<td>2.4</td>
</tr>
<tr>
<td>Malawi</td>
<td>0.400</td>
<td>446.2</td>
<td>59.6</td>
<td>64.3</td>
<td>1.7</td>
<td>39.1</td>
<td>1.0</td>
</tr>
<tr>
<td>Togo</td>
<td>0.495</td>
<td>1,555.8</td>
<td>178.6</td>
<td>55.0</td>
<td>3.3</td>
<td>52.2</td>
<td>3.1</td>
</tr>
<tr>
<td>Uganda</td>
<td>0.502</td>
<td>1,478.4</td>
<td>93.4</td>
<td>66.1</td>
<td>10.7</td>
<td>48.7</td>
<td>2.3</td>
</tr>
<tr>
<td>South Africa</td>
<td>0.653</td>
<td>11,192.2</td>
<td>1,917.3</td>
<td>27.6</td>
<td>6.6</td>
<td>47.0</td>
<td>1.6</td>
</tr>
<tr>
<td>Brazil</td>
<td>0.792</td>
<td>8,194.7</td>
<td>499.7</td>
<td>85.7</td>
<td>6.4</td>
<td>70.8</td>
<td>0.8</td>
</tr>
<tr>
<td>Mauritius</td>
<td>0.800</td>
<td>12,027.3</td>
<td>506.3</td>
<td>74.5</td>
<td>1.5</td>
<td>72.4</td>
<td>0.4</td>
</tr>
<tr>
<td>Costa Rica</td>
<td>0.841</td>
<td>9,481.4</td>
<td>1,016.6</td>
<td>72.4</td>
<td>1.2</td>
<td>78.3</td>
<td>0.5</td>
</tr>
<tr>
<td>Norway</td>
<td>0.955</td>
<td>35,453.5</td>
<td>4,225.7</td>
<td>100.0</td>
<td>0.9</td>
<td>80.6</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Notes: This Table is considerably shortened to nine countries only. The full set of countries is provided in Table 6 of the web-based version of this article (Wolff et al., 2010).
The HDI and Foreign Development Aid

Although, to our knowledge, the HDI is not formally used by any development agency as the sole index used to determine the distribution of development funds, there are clear indications that the HDI plays a significant role in governmental institutions’ and NGOs’ decisions for foreign aid allocation. In 2000, the Deputy Director of the UNDP exemplified this debate by stating:

‘At the global level, issues are now being explored as to whether bilateral aid can be allocated on the basis of HDI, or the core funds of multilateral agencies can be based on the index[...]’

(Jahan, 2000, p. 10).

In fact, ‘charity scorecards’ are increasingly used as a tool for helping individuals decide which countries to donate money to. Here, the HDI can be used to construct such a score. For example, on the homepage of http://www.charityscorecard.org/ (last accessed: 2 January, 2009) a world map of HDI scores is displayed. The use of the HDI in this context may explicitly and implicitly steer users to ‘misclassified countries’. Further, the triple bin classification is often used for report writing purposes to describe donor activities by governmental organizations (United Nations, 1996; HDR, 2001–2007) and non-governmental organizations. For example, Geneva Global (2007), which holds investments of 60 million client dollars in development projects, structures its funds according to the three HDI categories. For each year, the United Nations (HDR, 2001–2006) analyses the newest data on development aid as a function of the three human development categories. Drawing on these HDR statistics, Table 6 summarizes that across all years countries in the ‘low’ category obtained 3.4 times the official development assistance per capita as compared to the medium development countries, which we do not claim is a causal effect but rather an interesting correlation.

---

Use of the HDI Statistics in the Academic Literature

The HDI has been increasingly employed in the academic literature to describe the evolution of the world’s ‘welfare’ distribution in terms of various measures of inequality, such as the Gini coefficient, and to discuss the path of polarization, e.g. Pillarisetti (1997), Ogwang (2000), Mazumdar (2002), Noorbakhsh (2006), Prados de la Escosura (2007). The results published in these studies can differ greatly depending on which year the researcher collected the data in. To illustrate, Figure 9 displays HDI Gini coefficients using the formulas $h_A$, $h_B$ and $h_C$ for data covering 1975–2005 in five years intervals. The values produced by formula $h_A$ are 25 to 50% higher and the time trend steeper compared to the time series generated by formula $h_C$. This substantial difference would lead to different conclusions or policy recommendations by the analyst. For a recent discussion on the relevance of levels and gradients of Gini estimates see for example Sala-i-Martin (2006) and Prados de la Escosura (2007).
We find that a number of recent studies are sensitive to random selection of countries that is due to the ‘arbitrariness’ of the cut-off values: For example in the macroeconomic literature, Mazumdar (2002) and Noorbakhsh (2006) use the triple bins to analyze the existence of convergence clubs (Quah, 1996) by testing the beta and the sigma conditional convergence hypothesis, originally discussed in Barro and Sala-i- Martin (1992). In particular, Noorbakhsh (2006) runs beta-convergence regressions of the form

\[
\frac{\ln(hdi_{it+T}/hdi_{it})}{T} = \alpha + \beta \ln(hdi_{it}) + \epsilon_{it}
\]

conditional on the country belonging to the ‘low’ development bin. The dependent variable is the annualized growth of the HDI variable for country \(i\) over the period \(t\) to \(t + T\) and \(hdi_{it}\) is the ratio of HDI in the \(i\)th country to the average for the sample.\(^{15}\) The regression is then repeated for the bins ‘medium’ and ‘high’ and the comparison of the \(b\) estimates is used to analyze the existence of convergence clubs.

To illustrate the consequences of the random selection, we first rerun the convergence regression (3) conditional on the HDI being in the interval \(A_0 = [0.5, 0.8)\) as specified in Noorbakhsh (2006, p. 10, Table 3). Then we perform the same regression with the adjusted cut-off values in the set \(A_1 = [0.55, 0.70)\). The results are displayed in Table 7. Comparing the main parameter of interest, \(\beta\), the estimate of the second regression is about 100% off the first regression implying a much faster speed of convergence.\(^{16}\) This demonstrates that results based on the reported HDI can be very sensitive to changes of the HDI triple bin classification system.

\(^{15}\) A value of \(b\) in the range of \((-1, 0)\) would imply \(b\)-convergence of the countries in the sample. A \(\beta\) of zero means no convergence and a positive value for \(\beta\) indicates divergence, with the speed of convergence/divergence the higher the absolute value of \(\beta\).

\(^{16}\) Note that the two \(b\) estimates are statistically significant with \(t\) values of -6.74 and -4.59 for the sample of countries in \(A_0\) and in \(A_1\), respectively. We reject at the 1% significance level the hypothesis of uniform convergence in \(A_0\) and in \(A_1\) based on the Wald test examining whether \(\beta_1\) is different from \(\beta_0\), based on the pooled sample with appropriate interaction terms, with standard errors clustered by country.
Econometrically speaking, the average error measures $\hat{\sigma}_D$ and $\hat{\sigma}_F$ calculated in Section 3.1 imply that there is a 3 and 14\% downward attenuation bias in an ordinary least squares (OLS) regression, $y = \beta_1 + \beta_2 \text{HDI}^* + \varepsilon$, if the observed HDI – instead of the ‘true’ (but unknown) HDI$^*$ – is used as the regressor (for any variable $y$ of interest). The bias of the OLS estimate $b_2$ is given by\textsuperscript{17}

\[ \text{plim } b_2^D = \left[ \frac{1 - \sigma^2_D}{\sigma^2_D + \sigma^2_{\text{HDI}*}} \right] \beta_2 \approx 0.97\beta_2 \]

and

\[ \text{plim } b_2^F = \left[ \frac{1 - \sigma^2_F}{\sigma^2_F + \sigma^2_{\text{HDI}*}} \right] \beta_2 \approx 0.86\beta_2. \]

This is important as in many econometric cross-country studies the HDI is used as a regressor; see Globerman and Shapiro (2002), Mazumdar (2002), Sanyal and Samanta (2004), Neumayer (2003), Noorbakhsh (2006), Leigh and Wolfers (2006). This is even more crucial when working with the individual sub-indicator variables, since (as shown in Section 3.4) their average standard deviation of the measurement error is larger than the error of the HDI.

\textsuperscript{17}$\hat{\sigma}_D$ and $\hat{\sigma}_F$ is approximated by the empirical analogue of the 2006 HDI scores, $\hat{\sigma}_D \approx 0.036$. 

---

The figure shows the Gini Coefficients Computed by the HDI Formulas A, B and C.
Conclusions

This article identifies three sources of HDI data error and we make the following empirical contributions. First, we calculate country specific noise measures due to measurement error, formula choice and inconsistencies in the cut-off values. We find that the HDI statistics contain a substantial amount of noise on the order of 0.01–0.11 standard deviations. In analyzing the sources of the updating error we calculate country specific variances of GDP per capita, literacy rate, educational enrolment and life expectancy and we calculate the interdependence between these measures. We find that in general the higher the development status of a country, the more precise are the reported data. Second, we calculate the misclassification measures with respect to these three sources of data error by simulating the probabilities of being misclassified and sensitivity analysis of the cut-off values. We find that up to 45% of the developing countries are misclassified due to the failure to update the cutoff values. The discrete classification system is vulnerable when many countries are close to the thresholds, as is the case in the most recent years. Third, we discuss various empirical examples from the prior macroeconomic/development literature where the HDI has been employed and find that its use is problematic. Key parameters vary by up to 100% in their values. Although there may be certain benefits for the UN and charities for using a triple-bin classification system – bins are likely to improve publicity for the HDI and may hence help with more efficient internal organization of aid institutions – our results raise serious concerns about the system. We suggest that the United Nations should discontinue the practice of classifying countries into these triple bins because in our view the two cut-off values are arbitrary, can provide incentives for strategic behavior in reporting official statistics, and have the potential to misguide politicians, investors, charity donators and the public at large.

This article did not investigate the drivers of why in the early years of the HDI – when its political role was still uncertain – the distribution as displayed in Figure 1 looked so different from today’s. However, we caution governments, private investors, donor organizations and users of the charity scorecards not to take the triple bin system as a tool for international negotiations (Hu, 2009), foreign
direct investments (Arcelus et al., 2005), pricing (Bate and Boateng, 2007), or the allocation of foreign aid (Jahan, 2000; Neumayer, 2003). Such politically sensitive uses of the HDI might potentially provide perverse incentives for a country to manipulate the sub-indicator variables, if it has realized the comparative advantage of a 0.49 HDI score versus a 0.51 score. In fact, announcements such as the statement by Jahan (2000) (discussed in Section 4.2) might have just created these incentives. We quote Oskar Morgenstern (1970):

Governments, too are not free from falsifying statistics. This occurs, for example, when they are bargaining with other governments and wish to obtain strategic advantages or feel impelled to bluff [...]. A special study of these falsified, suppressed, and misrepresented government statistics is greatly needed and should be made.

References


