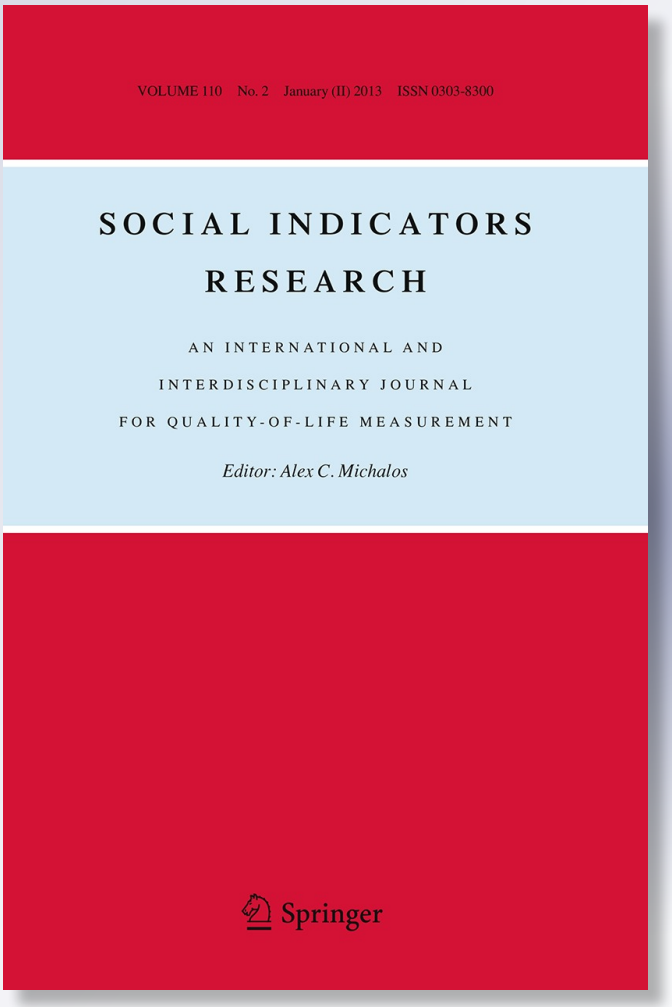
Monitoring Human Development Goals: A Straightforward (Bayesian) Methodology for Cross-National Indices

Kobi Abayomi & Gonzalo Pizarro

Social Indicators Research
An International and Interdisciplinary Journal for Quality-of-Life Measurement

ISSN 0303-8300
Volume 110
Number 2

Soc Indic Res (2013) 110:489-515
DOI 10.1007/s11205-011-9946-y



Springer

Soc Indic Res (2013) 110:489–515
DOI 10.1007/s11205-011-9946-y

Monitoring Human Development Goals: A Straightforward (Bayesian) Methodology for Cross-National Indices

Kobi Abayomi · Gonzalo Pizarro

Accepted: 3 October 2011 / Published online: 20 October 2011

Abstract We offer a straightforward framework for measurement of progress, across many dimensions, using cross-national social indices, which we classify as linear combinations of multivariate country level data onto a univariate score. We suggest a Bayesian approach which yields probabilistic (confidence type) intervals for the point estimates of country scores—a vital, and often missing, feature in cross-national comparisons. We demonstrate our approach using the United Nations Development Programme's Millennium Development Goals (MDGs), via the Maternal and Neonatal Program Effort Index (MNPI) data (Ross et al. in Trop Med Inter Health 6(10):787–798, 2001), and Human Development Index (HDI) (2010) as examples.

Keywords Millennium development goals · Indexing · Performance measurement · Bayesian statistics · Component analysis

1 Indexing

We call an index a metric—often constructed on administrative, spatial or heuristic units—that is used to characterize some salient, though latent—and perhaps not directly measurable—quality or quantity. For example: Gross Domestic Product (GDP) and the Dow Jones indexes are common economic indices; Pacific Decadal Oscillation (PDO) and El Nino (Francis et al. 1998; Gershunov and Barnett 1998) as climatological indices; the Human and Ecosystems Wellbeing Indexes—(HWI) and (EWI) (Prescott-Allen 2001) and

K. Abayomi (✉)
ISyE, Statistics, Georgia Institute of Technology, 765 Ferst Dr., 444 Groseclose, Atlanta, GA 30332, USA
e-mail: kobi@gatech.edu

G. Pizarro
MDG Support Team, Poverty Group, BDP UNDP, FF Building, 304 E 45th St. of 1052, New York, NY 10017, USA
e-mail: gonzalo.pizarro@undp.org

the United Nations Human Development Index—(HDI) (Place holder for Human Development 2011) are well known social indices.

Social indices seek to describe as well as predict phenomena that are often poorly measured and ill-defined. *A fortiori*, the act of constructing and reporting the index can yield new information, which can be used to guide more appropriate measurement or experimental design and refine future indexing (see Fuentes and A Holland 2006) for a creative example using Bernardo's 1979 fundamental comment on information maximization as a criteria).

Most indices, as functions on observed or observable data, are essentially linear or non-linear collections of (almost always) non-independent variables for the purpose of projecting a multidimensional concept onto a univariate scale of comparison. The scale of comparison—the range of the index—though arbitrary, is completely determined by the scheme for index construction and the characteristics of the underlying data (see for instance the Environmental Sustainability Index (ESI) (Abayomi et al. 2008, 2010). It is vital that any useful index be thoughtfully constructed in consideration of the way in which the consumers of the index—principally policymakers—typically focus on relative rankings rather than absolute scores. This is certainly true for development indices—where relative performance can drive international aid, excite or discourage potential donors, and (at least) bolster or embarrass politicians and elected officials.

1.1 An Index as a Statistical Object

Our goal in this paper is to suggest a straightforward framework for an index that remains a brief, cogent summary of important multidimensional concepts, accounts for measurement error, and conveys this information in a way that illustrates a discrimination among—or significant differences between—the results that policymakers will be able to use. Wolff et al. (see Wolff 2008) have illustrated the significant effect measurement error may have on an index score using the Human Development Index (see Place holder for Human Development 2011) as an example. By varying assumptions about the exactness of the data, the propriety of the computational formula, and the choice of quantile cut-offs for classifying countries they demonstrate a striking inconsistency with the reported values of the HDI.

Our contribution is consonant with Wolff et al's work in that we seek to incorporate Morgenstern's insistence (see Morgenstern 1970) on including distributional information (or variance) with point estimates. Our approach is a priori instead of *post hoc*, though, in that we offer a framework for the computation of measurement error available to the index constructor at the point of construction and not as a suffix or revision to completed work.

In the methodology section below we outline a generalized procedure for considering an observed value of a cross national index as some point estimate y generated as a linear combination of random predictors $\mathbf{X}$. We consider the construction of the index as an simultaneous estimation problem of the weights $\mathbf{c}$—the specific linear combination to use—the point estimate for each country and associated confidence intervals.

2 Methodology

Our approach is to define the multivariate data on which the index is defined as random variables with probability distributions. This assertion leads us to see the observed scalar index as a random quantity and as an estimate for some true characteristic. We consider the randomness, or error if you will, in the observed point estimate of the index (at each

country) to arise from the random distribution of the underlying data *and* the particular linear combination—choice of weights—used.

Intellectually we can consider the random distributions for the multivariate data as the *sampling* model for the index; we should consider a random distribution for the weighting scheme as the *design* model for the index. Below we consider the explicit consideration of both sampling and design randomness on probabilistic intervals for the country specific estimates. We in no way consider these illustrations definitive or complete, rather we suggest these a framework for understanding the eventual country scores as random objects with error bars around them. As well, we do not consider the area specific theoretical issues that may guide index constructors to select which, what and how to measure (see OECD 2008). Our methodology addresses the index specifically as an estimator of a univariate parameter which is the mapping of a multidimensional country level conceptual model to a univariate value. The choice of weights, the design, of course fixes a particular conceptual model—we address this below as a statistical issue and not more fully as an exogenously philosophical one.

2.1 Data

The data arrive in this methodology as

$$\mathbf{X} = (X_1, \ldots, X_k) \sim f_{\mathbf{X}}$$

a collection of ratings/scores with some multivariate, non-independent, distribution $f_{\mathbf{X}}$. Each X_j can be an 'average' from judges (say $1, \ldots, n_j$)—or not. Our focus here is the specification of y_i as a random score for country i, with an associated confidence interval (CI) for country i of the form:

$$\mathbb{P}(y_i \in (L_i, U_i)) = 1 - \alpha$$

with L_i and U_i the confidence bounds for each score. Thus we need a framework that generates a different CI for each y_i, i.e. for each country i; each y_i is a 'weighted score' of judge ratings on variables/items X_1 through X_k. That is:

$$y_i = \sum_{j=1}^{K} c_j X_j \tag{1}$$

The vector $\mathbf{c}^T$ is the 'weighting' scheme chosen for the index: Under the assumption that this scheme is constant across countries $i = 1, \ldots, N$, the CI's (at each country i) should then be a function of the randomness of a particular choice of scheme $\mathbf{c}^T$ as well as the distributional or sampling assumptions from the data $\mathbf{X}$.

Let $\mu = (\mu_1, \ldots, \mu_K)$ be the vector of means for the variables $\mathbf{X}$ in the index. Let $\sigma^2 = (\sigma_1^2, \ldots, \sigma^2{}_K)$ be the vector of variances. Notate $\sigma_{j,l} \equiv Cov(X_j, X_l)$ and collect the variances and covariances in the matrix Σ. The correlation is $\rho_{j,l} = \frac{\sigma_{j,l}}{\sigma_j \cdot \sigma_l}$; collect the correlations as $\rho = ((\rho_{j,l}))_{j<l=1\ldots K}$.

2.2 Confidence Intervals

Prevailing, comparable indexes lack proper probability or sampling models: country level scores in absence of distributional assumptions may be ordered and ranked—*but only in ignorance of statistically significant difference*.

The Human Development Index (HDI) (Human Development2009) and Environmental Sustainability Index (ESI) (World Economic 2001, 2002), for example, take opposite

approaches to modeling complexity: the HDI is an immediate combination of a small number of variables while the ESI is a weighted linear combination of many data sources. Neither of these indexes, though, yields any information on significance of differences in score (see also Adler et al. 2009).

In practice this leaves policy makers and stakeholders to compare magnitudes or rankings in obscurity of the sensitivity of the index to differential inputs. *A fortiori*, real differences between country effort are indistinguishable and unidentifiable. This flaw has severe implications and impacts: countries with truly differing scores may look similar, countries with similar scores may be judged identical—each error masking processes that need to be improved.

Three possible methods of generating the country-wise confidence intervals are:

- Distribution Free—minimal assumptions are placed on multivariate distribution of the Judges' ratings.
- Frequentist—Distributional assumptions on $f_{\mathbf{X}}$, the multivariate distribution of $\mathbf{X}$.
- Bayesian—Prior distributions on the parameterization of $f_{\mathbf{X}}$.

These approaches are listed in order of the restrictiveness of a priori assumptions: distribution free (distribution invariant) approaches impose the least assumptions on the data—the Bayesian approaches impose the most structure. Generally, a more definite model, one which requires stronger assumption, yields tighter confidence intervals for the parameter estimates.

2.2.1 'Distribution Free' Approach

For example, using the well known Tchebyshev's inequality we can write a 'distribution free' confidence interval (given known covariance matrix Σ as)

$$(1-\alpha) \equiv \mathbb{P}\left(y \in \sum_{j=1}^{K} c_j \bar{x}_j \pm t \right) \geq 1 - \frac{\sum_{i=1}^{K} c_j^2 \sigma_j^2 + 2\sum_{j<l} c_j c_l \sigma_{j,l}}{t^2} \tag{2}$$

which sets the $(1-\alpha)$ CIs to be $L_i \leq \sum_{j=1}^{K} c_j \bar{x}_j - t$ and $U_i \geq \sum_{j=1}^{K} c_j \bar{x}_j + t$.

2.2.2 'Simple Frequentist' Approach

Alternately we could suppose the joint distribution for the judges ratings is multivariate normal:

$$\mathbf{X} \sim f_{\mathbf{X}} \equiv N_K(\mu^T, \Sigma)$$

with univariate normal distributions that are identical across countries i, $\mu_{ij} = \mu_j$

$$X_{ij} \sim N(\mu_j, \sigma_j^2)$$

The expectation and variance of $y_i = \sum_{j=1}^{K} c_j X_j$ are as above:

$$E(y) = \sum_{j=1}^{K} c_j E(X_j)$$

$$Var(y) = \sum_{j=1}^{K} \sum_{j=1}^{K} c_j^2 \sigma_j^2 + 2 \sum_{j<l} c_j c_l \sigma_{j,l}$$

y_i then is distributed univariate normal since linear transforms of normal distributions are normally distributed.

$$y_i \sim N\left(\sum_{j=1}^{K} c_j\mu_j, \sum_{i=1}^{K} c_j^2\sigma_j^2 + 2\sum_{j<l} c_j c_l \sigma_{j,l}\right)$$

and the $(1 - \alpha)$ confidence interval for any y_i is

$$(1-\alpha) \equiv \mathbb{P}\left(y \in \sum_{j=1}^{K} c_j\mu_j \pm Z_{\alpha/2} \cdot \left(\sum_{i=1}^{K} c_j^2\sigma_j^2 + 2\sum_{j<l} c_j c_l \sigma_{j,l}\right)\right) \tag{3}$$

setting $(L_i, U_i) = \sum_{j=1}^{K} c_j\mu_j \pm Z_{\alpha/2} \cdot \left(\sum_{i=1}^{K} c_j^2\sigma_j^2 + 2\sum_{j<l} c_j c_l \sigma_{j,l}\right)$. These are fixed width CI's; contrast with the above distribution-free result where the CI width is slack and we take the most conservative bound.

We suggest and illustrate below what could be called a simple or naïve Bayesian approach in this paper: we fix the prior distributions to be conditionally independent and we initialize them with simple, exogenous estimates we can generate immediately. This is a commonly used approach on many types of data, straightforward, and flexible for different settings. See Gelman et al. (2004) for a good reference on the Bayesian approach to data modeling.

2.3 'Straightforward Bayesian' Framework

The Bayesian approach is to incorporate distributional assumptions on the parameters of interest. In this setting these parameters are introduced to yield *posterior* probability distributions for the country scores y_i and to impose *prior* probability distributions for the mean, covariance, and weighting parameters—μ^T, Σ and $\mathbf{c}^T$.

2.3.1 Multivariate Normal: Σ 'known'

Consider the case when the covariance matrix for $\mathbf{X}$ is known or (very) well estimated. The prior distribution

$$\mu^T \sim N(\mu_0^T, \Lambda_0)$$

assumes that the means are multivariate normal with μ_0^T, Λ_0 fixed (i.e. estimated from data). The *posterior distribution* for μ^T is

$$\pi(\mu^T|\mathbf{x}, \Sigma) \equiv N(\mu_n, \Lambda_n)$$

where

$$\mu_n = (\Lambda_0^{-1} + n\Sigma^{-1})^{-1}(\Lambda_0^{-1}\mu_0 + n\Sigma^{-1}\overline{\mathbf{x}})$$

and

$$\Lambda_n^{-1} = \Lambda_0^{-1} + n\Sigma^{-1}$$

Here $\mathbf{x}$ are the n observed judge ratings. Note that y is merely a linear transform of $\mathbf{x}$, in vector notation: $y = \mathbf{c}^T\mathbf{X}$. Thus y is univariate normal with

$$E(y) = \mathbf{c}^T \mu_n$$

and

$$Var(y) = Var(\mathbf{c}^T \mu_n) = \mathbf{c}^T \Lambda_n \mathbf{c}$$

The Bayesian CI's (often called Credible Intervals) are the random draws from the distribution; the posterior distribution here is multivariate normal. In this case we have closed form expressions for the expectation and variance of y—a reasonable approximate Bayesian CI is

$$(1 - \alpha) \equiv \mathbb{P}(y \in \mathbf{c}^T \mu_n \pm Z_{\alpha/2} \cdot \mathbf{c}^T \Lambda_n \mathbf{c}) \tag{4}$$

2.3.2 Multivariate Normal: Σ 'unknown'

The results are similar with the additional relaxation of a prior on the variance-covariance matrix Σ as well. A common prior is:

$$\Sigma \sim Inv - Wishart_{v_0}(\Lambda_0^{-1})$$

and

$$\mu | \Sigma \sim N(\mu_0, \Sigma / \kappa_0)$$

where v_0 and κ_0 are the degrees of freedom and scale matrix for the inverse-Wishart distribution on Σ. The joint posterior is multivariate normal. Sampling from the joint posterior to generate CI's for y can follow this algorithm (Gelman et al. 2004):

1. Draw $\Sigma | \mathbf{x} \sim Inv - Wishart_{v_0+n}(\Lambda_n^{-1})$
2. Draw $\mu^T | \Sigma, \mathbf{x} \sim N(\mu_n, \Sigma / \kappa_n)$
3. Compute $y = \mathbf{c}^T \mu$

with v_0 a parameter for the Inverse Wishart distribution. This yields a sampling posterior for y and the CI can be gleaned directly from inspection of the simulated replicates.

2.4 Considering the Weighting

Choosing the appropriate weighting scheme and generating CI's for each scalar y_i are separable tasks. The CI's are of course affected by the choice of weighting scheme, however, the weights themselves are arbitrary in the sense that they are subject to an exogenous constraint chosen by the indexers.

Desirable conditions on the choices on the weights could be:

- Maximal independence within $\mathbf{X}$
- Minimum covariance between X_i and X_j
- Maximum variation across scores y_i

2.4.1 Maximal Independence

Consider a model

$$\mathbf{Y} = B\mathbf{X}$$

where the components of $\mathbf{Y}$ are independent, and B is an estimate of A^{-1}, a mixing matrix for the latent/unobserved model:

$$\mathbf{X} = A\mathbf{S}$$

with $S \sim Q = \prod_{i=1}^{K} Q_i$. This is the Independent Component Analysis (ICA) model and algorithms exist to estimate B and thus the y as $\hat{\mathbf{S}}$.

Consider a diagonalization of B

$$B = \mathbf{L}^T D \mathbf{L}$$

with L an upper triangular matrix, and D a diagonal matrix. D yields a weighting scheme for the components of $\mathbf{X}$ and could be used as weights $\mathbf{c}^T$. Alternately, since $Y_j = B_j\,\mathbf{X}$—the 'independent' output of the ICA algorithm could be used as proxies for $\mathbf{X}$ in a null weighting scheme.

2.4.2 Minimum Covariance

Principal Component Analysis (PCA) can be viewed as a special case of the above ICA approach where Q is a multivariate Gaussian distribution (see Abayomi et al. 2008, 2010). The diagonalization of B is immediately

$$B = \Delta^T \mathbf{E} \Delta$$

where Δ and $\mathbf{E}$ are the eigenvectors and eigenvalues of the covariance matrix Σ in Q. Weighting items or components in this scheme is essentially Factor Analysis (Johnson and Wichern 1999).

2.4.3 Maximum Variation Across Scores

The output of the MDG indexing—a presentation of country-by-country scores (with confidence intervals and ranks)—suggests that maximizing variation across scores (across countries) is a desirable feature of a weighting scheme.

This goal may be addressed in a repeated measurement extension of the ICA or PCA algorithms, where the individual judge ratings are collected over all countries $\mathbf{X}_{i=1\ldots N}$

2.4.4 Bayesian Weighting

A direct approach is to let the $\mathbf{c}^T$ weights themselves have a prior distribution and investigate the distribution of y with this additional prioritization.

This is to model y as univariate normal as above:

$$y \sim N(\mathbf{c}^T \mu_n, \mathbf{c}^T \Lambda_n \mathbf{c})$$

with

$$\mu^T \sim N(\mu_0^T, \Lambda_0)$$

and

$$\Sigma \sim Inv - Wishart_{v_0}(\Lambda_0^{-1})$$

and

$$\mathbf{c}^T \sim Dirichlet(\alpha)$$

Sampling from the joint posterior to generate CI's for y can follow this algorithm:

1. Draw $\mathbf{c}^T \mid \mathbf{x} \sim Dirichlet(\alpha)$
2. Draw $\Sigma|\mathbf{x} \sim Inv - Wishart_{v_0+n}(\Lambda_n^{-1})$
3. Draw $\mu^T|\Sigma, \mathbf{x} \sim N(\mu_n, \Sigma/\kappa_n)$
4. Compute $y = \mathbf{c}^T \mu$

with $\alpha_1 = \cdots = \alpha_k = 1$; μ_n, κ_n and Λ_n^{-1} as before. In a Monte Carlo procedure this program is iterative and repeated until tolerance limits on the distribution of the parameters are satisfied. See Gelman and Hill (2006) for a fuller elucidation of this approach in varied settings.

We do note that this weighting approach is one of many possible: for example a naïve version of the Bayesian scheme here could be to set a degenerate distribution for $\mathbf{c}^T$ taking, for instance, each weight c_j as proportional to the sample variance of each X_j. This could be considered a straightforward frequentist approach.

The weighting scheme needn't be purely, or at all statistical. Hagerty et. al (2001) discusses varied weighting approaches for several extant indices; some rely not on past data but on prospective (prior) elicitation of expert opinion. In a strict sense this sort of divination, from expert opinion, can and should be framed as a statistical issue (see Gelfand et al. 1995); the point is that the weighting scheme is by no means necessarily derived from the variable predictors $\mathbf{x}$ or the index/response y.

Lastly, the weighting scheme presented here is particular to the class of indicators derived by linear (or perhaps log-linear) indexes. See Hagerty and Land (2007) for a more general discussion of weighting in the context of cross-administrative indices.

In the remainder of the paper we illustrate the Bayesian approach on the Human Development Index (2010) data and the Maternal and Neonatal Program Effort Index (MNPI) data (Ross et al. 2001), with relevance to the Millennium Development Goals (MDGs).

3 Maternal Mortality for the Millennium Development Goals

The United Nations Millennium Development Goals (MDGs) are eight objectives, by consent of the United Nations Member States in 2000, set out in the Millennium Declaration as benchmarks for reduction of poverty and hunger and increase of access to health care and education (MDG Task Force Progress 2010). Achievement of the Millennium Development Goals (MDGs) requires country-level, coordinated government efforts to reduce poverty and develop human resources, allied with efforts of private organizations and individuals (Millennium Development Goals 2010). These resources are realized financial, technical, and policy support from bilateral donors, multilateral institutions, and new sources of development finance such as philanthropic foundations (MDG Task Force 2010).

The existing monitoring of most of the elements of the MDG goals, operationalized in 21 specific targets and 60 indicators, is done systematically through the annual report on MDG progress (MDG Task Force 2010), which provides a comprehensive stocktaking across MDGs 1 through 7. Donor inputs, MDG 8, are tracked through the report of the MDG Gap Task Force (2010), which has become an annual publication. Other indices and

reports, such as the Commitment to Development Index (2009) and the annual ONE-DATA report (2009) on the fulfillment of commitments to Africa, also provide broad assessments of donor performance.

It is vital to model effort or performance "scores" for the MDGs as statistical, non deterministic objects. On the one hand, objective measures for distributional inequality are unlikely to be universally available (see Abayomi et al. 2008) and on the other much of the questionnaires are explicitly based on subjective expert ratings. The situation has some parallels to measurements for corruption, where objective measures are not readily available, particularly across countries. Early measures of corruption tended to be unreliable, being based on people's general impressions of the degree of corruption in a society. The weaknesses have been mitigated by carefully choosing respondents and designing questionnaires that focus on their actual experiences (Hawken 2007).

We illustrate as a first example of our methodology a country level index of progress toward the MDGs, specifically on the Maternal and Neonatal Health, MDG 5—reducing maternal mortality. Progress in this area has been measured previously across developing countries, using a reputation based approach, in the areas of family planning and, more recently, maternal and neonatal health and HIV/AIDS. In family planning, initial indicators were produced in 1972, using a questionnaire developed by leading analysts of family planning programs (Lapham and Mauldin 1972). Beginning with the second administration of the questionnaire in 1982, effort data were collected roughly every five years, and the seventh round of data collection is currently in progress.

We explicitly incorporate this via a repeated measures design and illustrate this component of the MDG index measurement. This approach is novel for this sort of data and in particular for MDG progress. (see Adler et al. 2009 for a non-probablistic contrast). We offer this example as a relatively sophisticated but directly implementable illustration.

3.1 Illustration: The MNPI Data

In 1999, a survey for maternal health with structure similar to the one we propose here was carried out in several countries as the Maternal and Neonatal Programme Effort Index (Bulatao and Ross 2002; Ross et al. 2001). The data contained in this survey provides an opportunity for testing and illustrating our proposed methodology. We offer a methodology that:

- illustrates issues that drive performance at a country level (i.e. discriminate the main drivers of variability, hence the weighting scheme needs to be appropriate and the same across countries).
- allows discrimination across countries (i.e. the methodology should be able to determine statistically significant index levels across countries).

The survey provides us with $N = 1{,}037$ observations by $K = 182$ variables: the judge ratings with metadata. The metadata are country and judge specific information. The rating data are variables 21–101—variables 102–182 are repeated measurements by each judge. These are the judge scores—$\mathbf{x}$—as outlined above. The metadata are variables 1–20 including country name and id. See the "Appendix".

3.2 Data Preparation: Imputation

The entire data (including the repeated measurements) have 9,505 missing values; 319 of the missing values are in the metadata for the judges. The percent of missing items is low

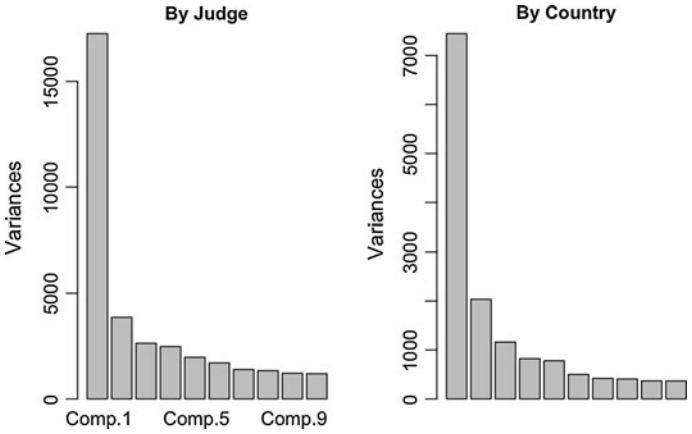


Fig. 1 *Scree* plots for variation of PCA by component. The *left graph* is the variation explained across judges, the *right* is across countries. A first component explains, respectively, 28 and 41% of the variation for each aggregation

(5%) but non-negligible. The location of the missing data, however, cannot be ignored. Missing data in both the meta-data and the covariates are imputed via *hot-deck*, this is, the completed data are re-samples of the observed at each country (see Little and Rubin 1987). A feature of the hot-deck procedure is that the model for the completions is explicitly empirical. The data were completed by hot-deck at each country to avoid collecting error beyond each set of country rankings.

The observations for Tanzania were discarded as many covariates were completely missing for all judges, thus reducing the total data to $N = 1022$.

To process the data and build the index, we **R** (The R Project for Statistical 2011), an free statistical programming language and open source versions of the Metropolis-Hastings algorithm (Williams 2001). We willingly provide sample code for our methodology upon request.

3.3 PCA for Null Weighting

Recall that the goal is to generate a score at each country which is a linear combination of the judge's ratings, $y_i = \sum_{j=1}^{K} c_j X_j$.

A priori, without any index or response variable to calibrate an initial or *null* weighting, a decision rule for the scheme can the desirable feature of minimal variance across rating items. In a sense, this is a projection of the collected rating items, the variables, to an orthogonal or independent basis. Weights assigned via a minimal variance scheme can identify (Gaussian or Normal) overdetermination in the covariates and suggest which may be discarded or of redundant importance in an index. See Bulatao and Ross (2002) for a prior, similar application (of factor analysis) to these data. See Fig. 1 for an illustration.

3.3.1 Aggregating Variation Across Judges

The PCA procedure (Sect. 3.2) is used to generate a set of null weights **c**. An initial PCA on the ungrouped data suggests the presence of some redundancy in the

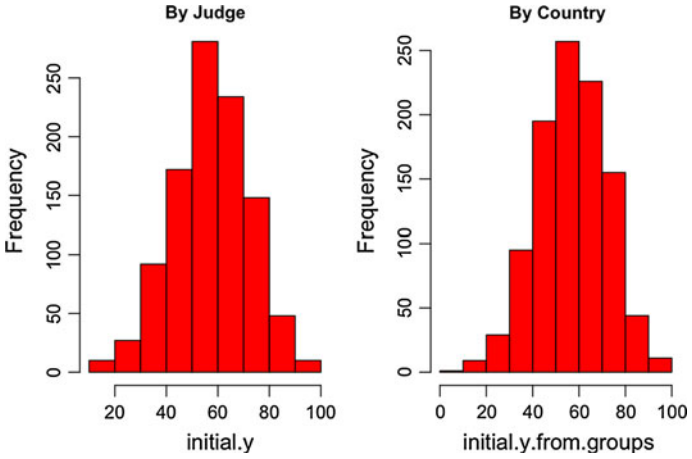


Fig. 2 Distribution of country scores, using PCA null weights, when aggregated by rater and by country. The maximum score by rater is Guajarat, by country is Jamaica

covariates; 28% of (Gaussian) variation can be explained by only one component, out of 81 possible.

The elements of the first eigenvector for the PCA decomposition are used as null weights: each $c_j \equiv e_j / \sum_j e_j$. Thus each $c_j \in (0, 1)$ and $\sum_j c_j = 1$.

This approach generates an index score for each judge, thus several for each country. The maximum score here was a judge rating for Gujarat and the minimum score was a for a rating of Yemen.

Null weighting by PCA when aggregated across judges may introduce inordinate bias to account for the variation within country, across judges. Notice that the maximum index score was generated by one (perhaps) optimistic rater for Yemen.

3.3.2 Aggregating Variation Across Countries

The PCA procedure under aggregation *across* countries estimates the eigenvectors—the null weights—via decomposition of the covariance matrix on the countries, instead of on the judges. This aggregation explains a higher proportion of the variation in the ratings, see Fig. 1. The maximum score—Jamaica; the minimum—Yemen (Fig. 2).

3.4 Bayesian Weighting

The scores generated by the PCA weighting are used as initial values in a Bayesian method for estimating the weights.

This is the scheme:

- Generate $\mathbf{c}_0^T$ as elements of first eigenvector from PCA. These null weights yield $y_0 = \mathbf{c}_0^T \mathbf{X}$, the null scores.
- Generate $Var(y_i) = \mathbf{c}_0^T \, Var(\mathbf{X}) \, \mathbf{c}$, the variance within a judges rating.
- Estimate $Var(y_0)$ as the sample variance of the null scores.

The PCA procedure provides the initial scores y_0 (generated from the null weighting scheme) and estimates for between and across variance.

- Let $y_i \sim N(\beta_g, \sigma_i)$, where the initial value of $\sigma_i = \sqrt{Var(y_i)}$. Here $i = 1 \ldots N$, the number of judges
- Let $\beta_g \sim N(\mathbf{c}^T \mathbf{X}, \sigma_g)$ be the country scores, where the initial value of σ_g is set to $\sqrt{Var(y_0)}$.
- Let $c_j \sim Dirichlet(\alpha)$ be the distribution for the weights. The initial weights are set identically to 1

This scheme allows a posterior to be estimated for β_g and c_j—the country specific scores and the variable weights. The posterior distributions yield confidence intervals for the country scores and the associated weights, automatically.

If all the judges ratings come from distributions with equivalent support—like {1, 2, 3, 4, 5} for Likert type or [0, 1] for percentages, say—the values of the weights can be interpreted as relative importance. The value of the weight for each item is the contribution of the item to the overall score, with respect to the way in which the weights are estimated.

In the example, the initial weights are assigned to maximize discrimination among countries; the resulting estimates are the relative contributions of items under this paradigm. These initial weights are starting estimates for the joint conditional estimation of the scores, weights, and associated variation.

Choosing a different weighting paradigm, via an alternate scheme, such as maximum variation among groups of countries or maximum inner product or score, yields different relative importances, of course, but with the same interpretation—modulo the method.

Of course, the weighting scheme may be adjusted to reconcile the judges responses, especially when the questions have nonequivalent support, such as some being "yes/no" items and others being rated {1, 2, 3, 4, 5}. The adjustment should leave the interpretation of the estimated weights unchanged.

Plots of the posterior distributions of the parameters for the country scores and variable weights are in Figs. 3 and 4.

4 The Human Development Index

The Human Development Index (HDI) was first introduced in 1990 by UNDP as a more comprehensive way to measure development as compared to income-based indicators, such as the GNP (Human Development 2009). The methodology has changed a bit over the life of the index (see Human Development 2010, 2011; Wolff 2008); in essence, and for the purpose of this illustration, the HDI is a weighted geometric mean of (sometimes rescaled) country level.

The 2010 HDI is

$$y_{HDI} = (X_{life} \cdot X_{edu} \cdot X_{GDP})^{1/3} \tag{5}$$

and we will generalize it with

$$HDI = (X_{life}^{c_1} \cdot X_{edu}^{c_2} \cdot X_{GDP}^{c_3}) \tag{6}$$

with $\sum_{j=1}^{k} c_j = 1$ as in the MDG example above. Equation 5 can be expressed

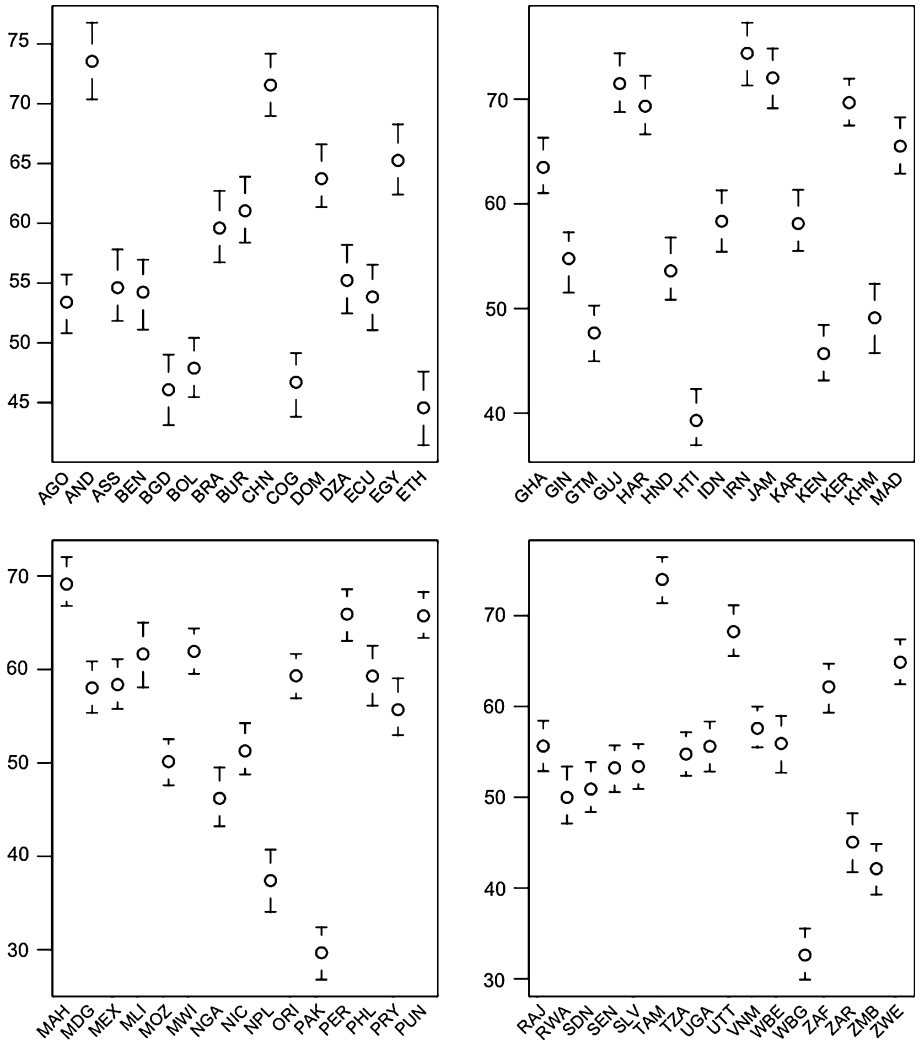


Fig. 3 Distribution of country scores, from posterior replicates, by alphabetical order of ISO3 country id code. The *upper* and *lower* 'whiskers' are the 75th and 25th percentiles of the posterior distribution

$$log(y_{HDI}) = c_1 log(X_{life}) + c_2 log(X_{edu}) + c_3 log(X_{GDP}) \tag{7}$$

which we can see as another version of Eq. 1, with $y \equiv log(y_{HDI})$ and $X_1 \equiv log(X_{life})$, etc. We used the publicly available data for the HDI which includes raw and rescaled values for life expectancy, literacy rate and gross domestic product for 135 countries from 1970 through 2010 (Human Development 2010).

4.1 Data Preparation

The publicly available HDI data set is complete for all years (three variables at each year) and all countries so there is no need to consider any imputation procedure. Wolff et al.

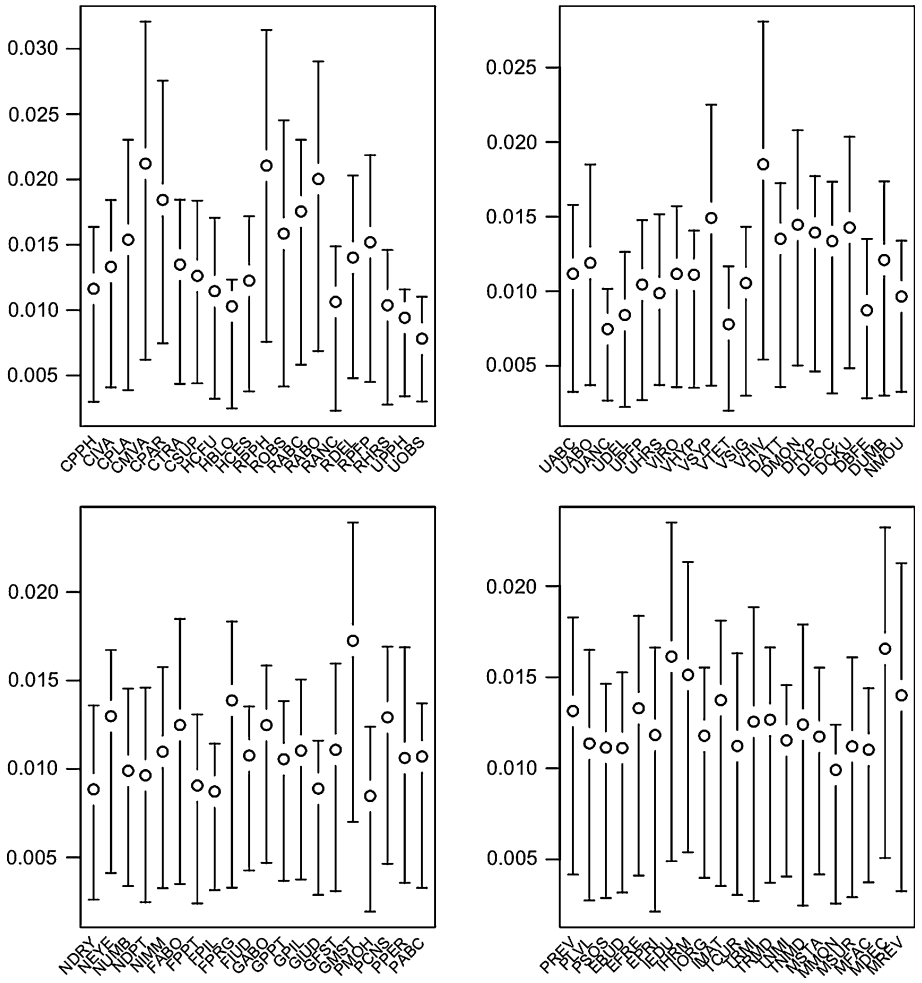


Fig. 4 Distribution of variable weights, from posterior replicates, by order of variable in questionnaire. The *upper* and *lower* 'whiskers' are the 75th and 25th percentiles of the posterior distribution

consider the effect of *post hoc* revisions of the measurements of the three HDI variables (life expectancy, literacy and GDP) and demonstrate appreciable randomness in HDI scores (Wolff 2008). We consider our example of an HDI with error bars to be a complementary illustration.

The HDI variables are rescaled versions of widely available life and income statistics over the 135 countries measured. For example: the life expectancy value X_{life} is the ratio of the difference between a country's observed, i.e. estimated, life expectancy at a given year and a minimal value (set at 20 years) to a maximal such difference—63 years: Japan's 83 years, observed in 2010, minus 20 (Human Development 2010). These choices are arbitrary and perhaps quite defensible; we do not address them as modeling issues here and focus on the rescaled and not the raw values.

We operate on the log transformed data as represented in Eq. 7, which allows us to remain in our linear setup, and exponentiate for graphs and illustrations

4.1.1 PCA for Null Weighting

Again we want to consider a choice for **c** driven by statistical methodology and we choose to initialize values under maximal variation across countries. Here the PCA program is to find the weighting assignment that maximizes variation across the 135 countries on three variables; the initialization weights we choose are the rescaled elements of the first eigenvector of the PCA decomposition. This yielded the initial weighting scheme in Table 1 below.

This initialization yields Australia with the maximum HDI score and Zimbabwe with the minimum. There are no repeated measurements at each year in the HDI data (i.e no multiple judge ratings as in the MDG MNPI example above)

4.2 Bayesian Weighting

Again we use the scores generated by the PCA weighting as initial values in a Bayesian estimation procedure for the weights and final scores.

Here is the scheme:

- Generate $\mathbf{c}_0^T$ as elements of first eigenvector from PCA. These null weights yield $y_0 = \mathbf{c}_0^T \mathbf{X}$, the null scores, in Fig. 5.

Table 1 Scaled first PCA eigenvector as initial weights for GDI

c_{life}	c_{edu}	c_{GDP}
0.23	0.21	0.56

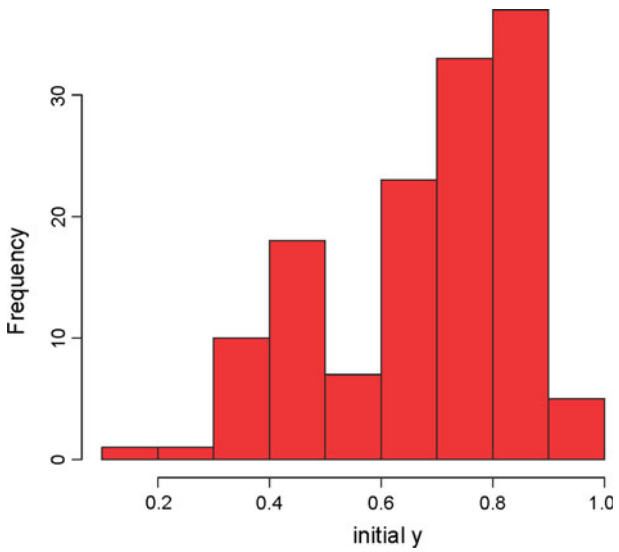


Fig. 5 Initial HDI scores with weightings set by first PCA eigenvector, for maximal index variation across countries

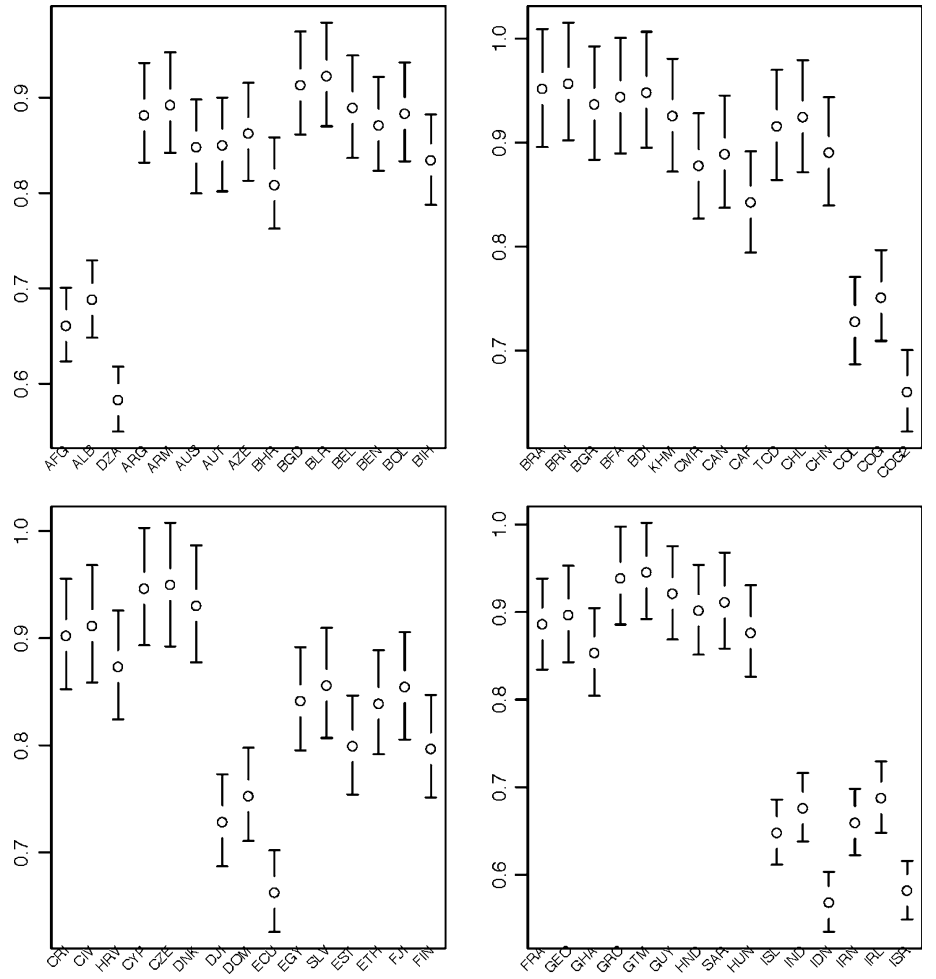


Fig. 6 Distribution of country scores, from posterior replicates, by alphabetical order of ISO3 country id code. The *upper* and *lower* 'whiskers' are the 97.5th and 2.5th percentiles of the posterior distribution

- Estimate $E(y_i)$ and $Var(y_i)$ with the sample mean and variance across all years (1970–2010) of each country's HDI score.
- Estimate $E(X_{ij})$ and $Var(X_{ij})$ as the sample mean and variance across all years (1970–2010) of each country's life, education and GDP values.

We incorporate these estimates in the Bayesian procedure

- Let $y_i \sim N(\mathbf{c}^T X_{ij}, \sigma_i)$, where the initial value of $\sigma_i = \sqrt{Var(y_i)}$. Here $i = 1 \ldots N$, the index over countries.
- Let $c_j \sim Dirichlet(\alpha)$ be the distribution for the weights. The initial weights are set identically to 1.
- Let $X_{ij} \sim N(\mu_{ij}, \sigma_{ij})$ where the μ_{ij} and σ_{ij} are estimated from the data record as above.

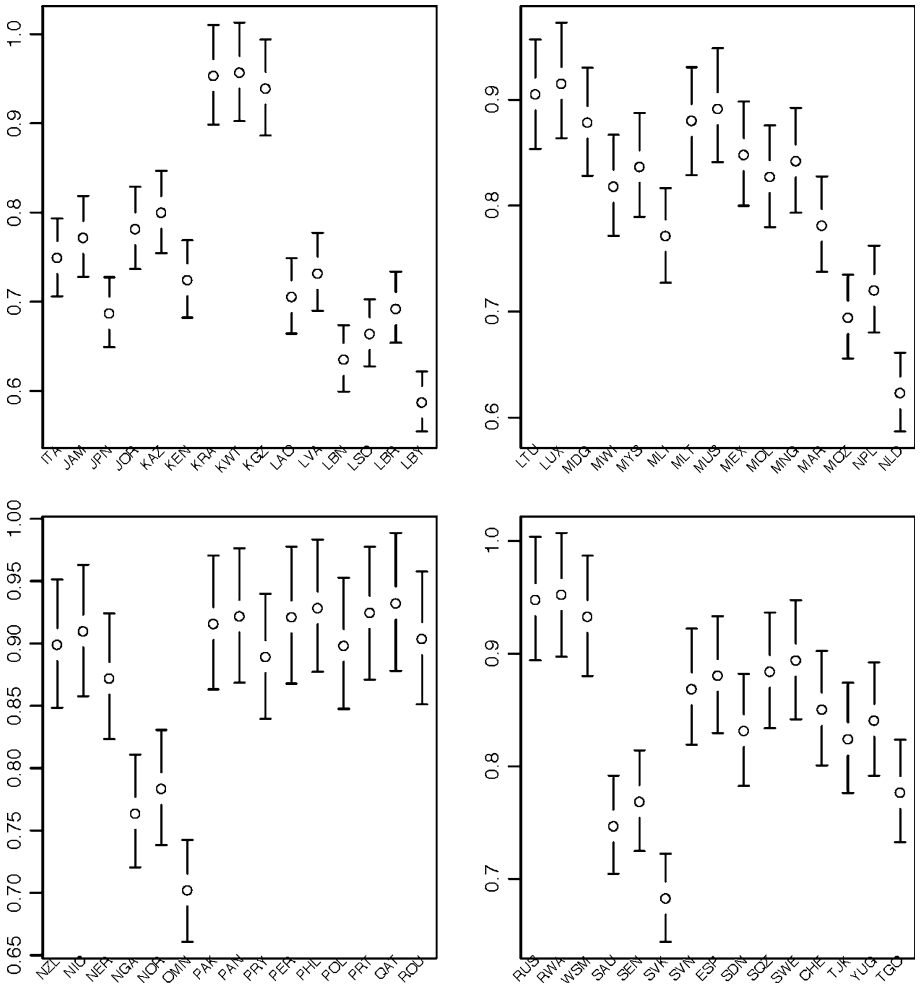


Fig. 7 Distribution of country scores, from posterior replicates, by alphabetical order of ISO3 country id code. The *upper* and *lower* 'whiskers' are the 97.5th and 2.5th percentiles of the posterior distribution

Similar to the above example this scheme allows a posterior to be estimated for y_i and c_j—the country specific scores and the variable weights. The posterior distributions yield confidence intervals for the country scores and the associated weights, automatically. The HDI scores are then back transformed via exponentiation to values on [0, 1]. See Figs. 6, 7 and 8.

As in the MDG-MNPI example above notice that many countries have scores that differ nominally but not statistically (for example Afghanistan and Albania in Fig. 6) which is the main point of the methodology. Contrast these illustrations with the point estimate rankings generate by the ordinary HDI (Human Development 2010): a practitioner would perhaps replace the ordering from 1…135 with (statistically) distinct ordered groups of statistically in-differentiable country scores.

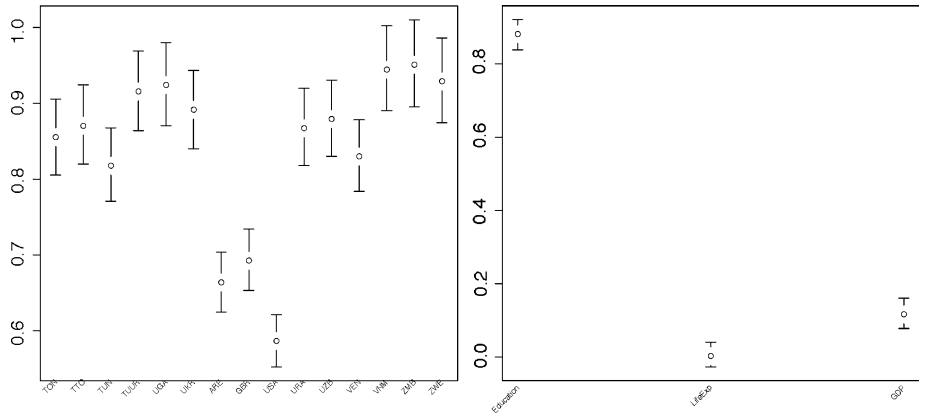


Fig. 8 *Left panel* Distribution of country scores, from posterior replicates, by alphabetical order of ISO3 country id code. The *upper* and *lower* 'whiskers' are the 97.5th and 2.5th percentiles of the posterior distribution. *Right panel* Distribution of weights, from posterior replicates, for HDI weights, 97.5th and 2.5th percentiles

5 Discussion and Summary

We have presented a framework for cross-national indices as statistical objects and demonstrated our approach on an indicator designed to measure for progress and effort toward the maternal health component of Millennium Development Goals (MDGs) and on the well known Human Development Index (HDI). Our methodology is designed to output not only point estimates of country level scores but probabilistic intervals for those estimates as well as for the weighting scheme that aggregates the variables the score is measured on. We used a Bayesian framework to generate these intervals by supposing prior distributions on the underlying data and variable weights and then examining the posterior replicates. We initialized simulations of these posterior replicates by supposing an initial weighting scheme—one of maximal variation across countries—using the well known Principal Component Analysis (PCA) procedure .

In the MNPI-MDG illustration we were able to 'borrow' inference from the repeated measurement design of the MNPI questionnaire (Bulatao and Ross 2002) and achieved relatively tight intervals (even at 50% confidence). The intervals the posterior replicates yield for the weights of the MNPI-MDG index are much wider at an equivalent level of confidence: perhaps mainly because of the high number of variables (questionnaire items) in the index.

In the HDI example we fixed the parameters of the prior distributions for the weights with the estimates of mean and variance from the time series of HDI data. The posterior replicates for the weights of the HDI index are all statistically different: the vast majority of the weight is assigned to the education variable in the HDI index.

We do not make any claim to the propriety of the examples offered here; in fact practitioners may choose very different paradigms in disagreement with our choices of prior distributions, principle of maximal variation, etc. Our contribution is to offer a method which allows for the comparison of countries in terms of statistically significant distance. Ranking and ordering point estimates without consideration of this distance exaggerates false differences, obscures possible policymaking levers and can. This methodology, which yields the significance of differences in country scores at a glance, can

accelerate and coordinate global responses especially for possible MDG process shortcomings. At the same time, the intervals for the weighting scheme yield an immediate picture of factors—including, perhaps: measurement error, rater bias, trends or change points—which affect the country scores.

We have focused particularly on human development indices in this paper, especially because the concepts practitioners and policymakers need to measure are more ethereal, perhaps, than in other settings. The Bayesian framework we offer is uniquely able to account for specificity or vagueness—as need be—via the prior distributions on weights and variables.

Appendix

Potential Questionnaires on MDG Goals and Targets

A list of 15 questionnaires is suggested to parallel, though not exactly duplicate, the lists of MDGs and targets. These are listed in Table 2, which shows the goals and targets to which each refers. It also shows the output indicators related to each questionnaire that have been proposed in other documents. These output indicators were meant to be suggestive rather than comprehensive, presumably chosen at least partly for the availability of reliable data. What the questionnaires should address is the effort that has gone or is going into improving not only these outputs but also other outputs related to the broader goals and targets. The list in Table 2 follows the order of the MDGs.

The MNPI Effort Questionnaire

A outline of a questionnaire on effort at achieving the maternal mortality target is provided here, by design of Ross et al. (2001). The data for the illustration in the paper follow this organization. We do not reproduce the entire questionnaire here.

Organization of the Questionnaire

The questionnaire is organized in two parts. The first, much longer part requests ratings of different features of a maternal health program. The second, short part (labeled "General background") requests relatively objective information about laws, plans, budgets, facilities, etc. relating to maternal health. All respondents are expected to answer the first part, but only a few, those more closely connected with the government maternal health program, are to be given the second part to answer. Though the two parts are somewhat different in format, they are not separated so that respondents who receive both parts will see them as a single questionnaire.

Substantively, the questionnaire covers typical project components of policy and planning, funding, service delivery, and demand generation. However, questions are not posed in this order, but start with service delivery. The purpose is to fix the respondent's attention initially on what services actually reach women in need and can have direct effect on reducing maternal mortality. The questionnaire seeks to emphasize what is actually making a difference on the ground rather than what agreements and plans are made on paper. After asking about services in several different ways, the questionnaire moves to more general policy issues.

Table 2 Proposed questionnaires on effort and the goals, targets, and output indicators they should cover

Questionnaire	Covers these goals and targets	Covers activities to affect these output indicators (among other possible ones)
Income, employment, and equity	*Goal 1*: Eradicate extreme poverty and hunger	*1.1* Proportion of population below $1 (PPP) per day
	Target 1.A: Halve, between 1990 and 2015, the proportion of people whose income is less than one dollar a day	*1.2* Poverty gap ratio
	Target 1.B: Achieve full and productive employment and decent work for all, including women and young people	*1.3* Share of poorest quintile in national consumption
		1.4 Growth rate of GDP per person employed
		1.5 Employment-to-population ratio
		1.6 Proportion of employed people living below $ 1 (PPP) per day
		1.7 Proportion of own-account and contributing family workers in total employment
		3.2 Share of women in wage employment in the non-agricultural sector
Food and nutrition	*Goal 1:* Eradicate extreme poverty and hunger	*1.8* Prevalence of underweight children under five years of age
	Target 1.C: Halve, between 1990 and 2015, the proportion of people who suffer from hunger	*1.9* Proportion of population below minimum level of dietary energy consumption
Education	*Goal 2:* Achieve universal primary education	*2.1* Net enrolment ratio in primary education
	Target 2.A: Ensure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling	*2.2* Proportion of pupils starting grade 1 who reach last grade of primary
		2.3 Literacy rate of 15–24 year-olds, women and men
	Target 3.A: Eliminate gender disparity in primary and secondary education, preferably by 2005, and in all levels of education no later than 2015	
		3.1 Ratios of girls to boys in primary, secondary and tertiary education

Table 2 continued

Questionnaire	Covers these goals and targets	Covers activities to affect these output indicators (among other possible ones)
		6.4 Ratio of school attendance of orphans to school attendance of non-orphans aged 10–14 years
Gender	*Goal 3:* Promote gender equality and empower women	*3.3* Proportion of seats held by women in national parliament
Child mortality	*Goal 4:* Reduce child mortality	*4.1* Under-five mortality rate
	Target 4.A: Reduce by two-thirds, between 1990 and 2015, the under-five mortality rate	*4.2* Infant mortality rate
		4.3 Proportion of 1-year-old children immunised against measles
Health Systems	*Goal 4, 5 and 6 Health system questionnaire* to be promoted to be used whenever any of the other health Questionnaires are used	
Maternal and neonatal health	*Goal 5:* Improve maternal health	*5.1* Maternal mortality ratio
	Target 5.A: Reduce by three quarters, between 1990 and 2015, the maternal mortality ratio	*5.2* Proportion of births attended by skilled health personnel
		5.5 Antenatal care coverage (at least one visit and at least four visits)
	Target 4.A: Reduce by two-thirds, between 1990 and 2015, the under-five mortality rate	*4.2* Infant mortality rate
Family planning	*Target 5.B:* Achieve, by 2015, universal access to reproductive health	*5.3* Contraceptive prevalence rate
		5.4 Adolescent birth rate
		5.6 Unmet need for family planning
HIV/AIDS	*Goal 6:* Combat HIV/AIDS, malaria and other diseases	*6.1* HIV prevalence among population aged 15–24 years
		6.2 Condom use at last high-risk sex
	Target 6.A: Have halted by 2015 and begun to reverse the spread of HIV/AIDS	*6.3* Proportion of population aged 15–24 years with comprehensive correct knowledge of HIV/AIDS
	Target 6.B: Achieve, by 2010, universal access to treatment for HIV/AIDS for all those who need it	*6.5* Proportion of population with advanced HIV infection with access to antiretroviral drugs

Table 2 continued

Questionnaire	Covers these goals and targets	Covers activities to affect these output indicators (among other possible ones)
Major parasitic and infectious diseases	*Goal 6:* Combat HIV/AIDS, malaria and other diseases	*6.6* Incidence and death rates associated with malaria
	Target 6 .C: Have halted by 2015 and begun to reverse the incidence of malaria and other major diseases	*6.7* Proportion of children under 5 sleeping under insecticide-treated bednets
	Target 8.E: In cooperation with pharmaceutical companies, provide access to affordable essential drugs in developing countries	*6.8* Proportion of children under 5 with fever who are treated with appropriate anti-malarial drugs
		6.9 Incidence, prevalence and death rates associated with tuberculosis
		6.10 Proportion of tuberculosis cases detected and cured under directly observed treatment short course
		8.13 Proportion of population with access to affordable essential drugs on a sustainable basis
Environment	*Goal 7:* Ensure environmental sustainability	*7.1* Proportion of land area covered by forest
	Target 7 .A: Integrate the principles of sustainable development into country policies and programmes and reverse the loss of environmental resources	*7.2* $C0_2$ emissions, total, per capita and per $1 GDP (PPP)
	Target 7.B: Reduce biodiversity loss, achieving, by 2010, a significant reduction in the rate of loss	*7.3* Consumption of ozone-depleting substances
		7.4 Proportion offish stocks within safe biological limits
		7.5 Proportion of total water resources used
		7.6 Proportion of terrestrial and marine areas protected
		7.7 Proportion of species threatened with extinction
Water and sanitation	*Target 7 .C:* Halve, by 2015, the proportion of people without sustainable access to safe drinking water and basic sanitation	*7.8* Proportion of population using an improved drinking water source
		7.9 Proportion of population using an improved sanitation facility

Table 2 continued

Questionnaire	Covers these goals and targets	Covers activities to affect these output indicators (among other possible ones)
Urban slum improvement	*Target 7.D:* By 2020, to have achieved a significant improvement in the lives of at least 100 million slum dwellers	*7.10* Proportion of urban population living in slums
Communication	*Target 8.F:* In cooperation with the private sector, make available the benefits of new technologies, especially information and communications	*8.14* Telephone lines per 100 population
		8.15 Cellular subscribers per 100 population
		8.16 Internet users per 100 population
Global partnership	*Goal 8:* Develop a global partnership for development	*Official development assistance (ODA)*
	Target 8 .A: Develop further an open, rule-based, predictable, non-discriminatory trading and financial system. Includes a commitment to good governance, development and poverty reduction—both nationally and internationally	*8.1* Net ODA, total and to the least developed countries, as percentage of OECD/DAC donors' gross national income
	Target 8.B: Address the special needs of the least developed countries, including: tariff and quota free access for the least developed countries' exports; enhanced programme of debt relief for heavily indebted poor countries (HIPC) and cancellation of official bilateral debt; and more generous ODA for countries committed to poverty reduction	*8.2* Proportion of total bilateral, sector-allocable ODA of OECD/DAC donors to basic social services (basic education, primary health care, nutrition, safe water and sanitation)
	Target 8 .C: Address the special needs of landlocked developing countries and small island developing States (through the Programme of Action for the Sustainable Development of Small Island Developing States and the outcome of the twenty-second special session of the General Assembly)	*8.3* Proportion of bilateral official development assistance of OECD/DAC donors that is untied

Table 2 continued

Questionnaire	Covers these goals and targets	Covers activities to affect these output indicators (among other possible ones)
	Target 8.D: Deal comprehensively with the debt problems of developing countries through national and international measures in order to make debt sustainable in the long term	*8.4* ODA received in landlocked developing countries as a proportion of their gross national incomes
		8.5 ODA received in small island developing States as a proportion of their gross national incomes
		Market access
		8.6 Proportion of total developed country imports (by value and excluding arms) from developing countries and least developed countries admitted free of duty
		8.7 Average tariffs imposed by developed countries on agricultural products and textiles and clothing from developing countries
		8.8 Agricultural support estimate for OECD countries as a percentage of their gross domestic product
		8.9 Proportion of ODA provided to help build trade capacity *Debt sustainability*
		8.10 Total number of countries that have reached their HIPC decision points and number that have reached their HIPC completion points (cumulative)
		8.11 Debt relief committed under HIPC and MDRI Initiatives
		8.12 Debt service as a percentage of exports of goods and services

Table 3 Classification of questionnaire items

Category	Questionnaire item numbers
1. Policy and planning: Policy is taken here in the sense of laws, regulations, standards, and guidelines that affect individual behavior relating to maternal health, the functioning of the maternal health program, and the conduct of service providers and others with whom they must interact. Plans are mainly national plans	
1.1 Appropriate laws	59, 98, 99, 100a
1.2 Regulations and guidelines	56–60, 100
1.3 Plans	65, 71–72, 101–103
2. Budget and finance: Government budgets are covered as well as financing. Because cost recovery is discouraged in regard to maternal health services, local finance comes mainly from sources outside the health sector and is not specifically covered	
2.1 Budget and expenditures	66–67, 69, 71, 74–76, 106–108
2.2 Donor support	68, 109, 112–116
2.3 Harmonization of activities	70, 110–111, 117
3. Service delivery: Different aspects of effective service delivery are listed below. Services usually require all of these elements to succeed, so most of the items could fall under most of the headings. However, each relevant questionnaire item is generally listed only under quality services and one other heading, reflecting its major emphases	
3.1 Quality services	1–55
3.2 Adequate facilities	1–12, 20–21, 36, 72, 92, 95, 97, 105
3.3 Competent staffing	1–6, 10–18, 22–35, 49, 53, 78–86, 92, 104
3.4 Appropriate supplies and equipment	7–8, 47–48, 52, 73, 91
3.5 Equitable attention	9, 13–2l,[a] 36–38, 74, 90, 65a, 65b, 86a, 97a
3.6 Effective monitoring and evaluation	60, 92–97
4. Demand generation: This involves mobilizing social groups and communities and providing good information to women and households about what needs to be done to avoid maternal deaths	
4.1 Information, education, communication	62, 87–88, 91, 91a
4.2 Social mobilization	57, 89–90
5. Governance: Good governance cuts across the preceding categories requiring everything from sound policy to effective government services. One dimension of good governance as defined by the World Bank Institute, political stability and the absence of terrorism, is not directly assessed in its impact on the sector, and is left out here. A second dimension, government effectiveness, practically covers all the items, so it is represented by the more limited category of an effective management structure	
5.1 Voice	57, 90, 97b, 97c
5.2 Effective management structure	61, 63–64
5.3 Regulatory quality	77
5.4 Rule of law	59[b]
5.5 Control of corruption	69, 76

[a] These items address rural–urban differentials

[b] There is an international agreement that post-abortion care should be provided. "Disregarding the law" questions, however, do suggest that some assessment of (de)criminalization and stigma need to be assessed

Questions are not necessarily grouped in categories familiar to donors. Instead, they are grouped for convenience, keeping together those with a similar frame of reference requiring answers in a similar format. Nor are questions intended as a checklist of all the specific requirements for providing proper maternal care. To keep the questionnaire at reasonable length and to avoid asking about details too fine for some respondents, the questions necessarily reflect a sampling of important best practices and dimensions of

effort. To indicate how responses might be reclassified, after the data are obtained, to reflect particular issues of relevance from a planning perspective, Table 3 provides an illustration The table lists some items more than once, as reflecting more than one aspect of performance. Some items could be listed under even more categories. Subsequent empirical analysis may suggest the most useful groupings.

References

Abayomi, K., Gelman, A., & Levy, M. (2008). Diagnostics for multivariate imputation. *Journal of the Royal Statistical Society-C, 57*(Part 3), 1–19.

Abayomi, K., de la Pena, V., & Lall, U. (2008). *Copula based independent component analysis (CICA)*. Working paper.

Abayomi, K., de la Pena, V., Lall, U., & Levy, M. (2010). Quantifying sustainability: Methodology for and determinants of an environmental sustainability index chapter in *Green Finance and Sustainability*. IGI Global.

Adler, N., Yazhemsky, E., & Tarverdyan, R. (2009). A framework to measure the relative socio-economic performance of developing countries. *Socio-Economic Planning Sciences, 3*, 1–16.

Bernardo, J. (1979). Expected information as expected utility. *Annals of Statistics, 7*(3), 686–690.

Bulatao, R. A., Ross, J. A. (2002). Rating maternal and neonatal health services in developing countries. *Bulletin of the World Health Organization, 80*, 721–727.

Commitment to Development Index. (2009). *Center for global development*. Washington, DC. Accessed at http://www.cgdev.org/section/initiatives/_active/cdi/ on 20 October 2009.

Francis, R. C., Hare, S. R., Hollowed, A. B., & Wooster, W. S. (1998). Effects of interdecadal climate variability on the oceanic ecosystems of the Northeast Pacific. *Fisheries Oceanography, 7*, 22.

Fuentes, M. C., & A Holland, D. (2006). Bayesian entropy for spatial sampling design of environmental data. *Environmental and Ecological Statistics*. June 2006.

Gelfand, A., Mallick, B., & Dey, D. (1995). Modeling expert opinion arising as a partial probabilistic specification. *Journal of the American Statistical Association*. 90, 430 (June 1995).

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian data analysis* (2nd Ed.). London: Chapman Hall/CRC.

Gershunov, A., & Barnett, T. P. (1998). Interdecadal modulation of ENSO tele-connections. *Bulletin of the American Meteorological Society, 79*, 12.

Hagerty, M., & Land K. (2007). Constructing summary indices of quality of life—A model for the effect of heterogenous importance weights. *Sociological Methods & Research, 55*(1), 455–496.

Hagerty, M., Cummins, R., Ferriss, A., et. al (2001). Quality of Life Indexes for national policy: Review and agenda for research. *Social Indicators Research, 55*, 1–96.

Hawken, A., & Munck, G. L. (2007). Measuring corruption: A critical assessment and a proposal. *Technical Background Paper for Asia Pacific Human Development Report*. UNDP.

Human Development Report 2009. (2009). United Nations Development Programme.

Human Development Report 2010. (2010). United Nations Development Programme.

Human Development Report. (2011). UNDP. Oxford University Press.

Johnson, R., & Wichern, D. (1999). *Applied multivariate statistical analysis*. London: Prentice Hall.

Lapham, R. J., & Mauldin, W. P. (1972). National family planning programs: Review and evaluation. *Studies in Family Planning, 3*(3), 29–52.

Little, R., & Rubin, D. (1987). *Statistical Analysis with Missing Data*. London: Wiley.

MDG Task Force Progress Chart. (2010). New York. Accessed at http://unstats.un.org/unsd/mdg/Resources/Static/Products/Progress2010/MDG_Report_2010_Progress_Chart_En.pdf

MDG Task Force Report. (2010). New York. Accessed at http://mdgs.un.org/unsd/mdg/Resources/Static/Products/Progress2010/MDG_Report_2010_En.pdf

Millennium Development Goals Indicators. (2010). The official United Nations site for the MDG Indicators. Available at http://unstats.un.org/unsd/mdg/Host.aspx?Content=Indicators/OfficialList.htm

Morgenstern, O. (1970). *On the accuracy of economic observations* (2nd ed.). Princeton: Princeton University Press.

OECD. (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Paris: OECD and European Commission. OECD Publishing.

Prescott-Allen, R. (2001) *The wellbeing of nations: a country-by-county index of quality of life and the environment*. Washington, D.C: Island Press/The Center for Resource Economics.

Ross, J. A., Campbell, O. M. R., & Bulatao, R. (2001). The maternal and neonatal programme effort index (MNPI). *Tropical Medicine and Internal Health, 6*(10), 787–798.

Stover, J. (1999). *The AIDS programme effort index (API): Results from the field test*. Washington, DC: Futures Group.

The Data Report 2009. (2009). Monitoring the G8 Promise to Africa, 19 May 2009. Accessed at http://www.one.org/international/datareport2009/pdfs/DR2009.pdf on 20 October 2009.

The Millennium Development Goals Report. (2009). The United Nations Development Programme. New York.

The R Project for Statistical Computing. (2011). Available at http://www.r-project.org/.

Williams, D. (2001). *Weighing the odds: A course in probability and statistics*. Cambridge.

Wolff, H., Chong, H., & Auffhammer, M. (2008) Consequences of data error in aggregate indicators: Evidence from the human development index. *Report* Department of Agricultural and Resource Economics. UC Berkeley.

World Economic Forum. (2001). Environmental sustainability index. Global leaders for tomorrow environment task force. World Economic Forum and Yale Center for Environmental Law and Policy and Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network. Davos, Switzerland and New York. Available at: http://sedac.ciesin.columbia.edu/es/esi/.

World Economic Forum. (2002). Environmental sustainability index. Global leaders for tomorrow environment task force. World Economic Forum and Yale Center for Environmental Law and Policy and Yale Center for Environmental Law and Policy and Center for International Earth Science Information Network. Davos, Switzerland and New York. Available at: http://sedac.ciesin.columbia.edu/es/esi/.