DEPARTMENT OF STATISTICS
Uppsala University
Bachelor Thesis
Spring 2011
Supervisor: Lennart Norell
Authors: Cecilia Lindman and Jenny Sellin

Measuring Human Development

The Use of Principal Component Analysis in Creating an Environmental Index

Abstract

In the current development debate, much critique has been directed towards the modern measures of welfare that attempt to assess human well-being according to more dimensions than just per capita income. One such measure is the United Nations Development Programme's Human Development Index (HDI), which is comprised by measures of education, health and income. The HDI has been subject to many evaluations concluding that it is intrinsically correlated with GDP, and is believed to possess poor statistical quality due to the arbitrary weighting of the included variables. It has also been proposed that the environment and its sustainability are of importance to human welfare, and should be included in measures of well-being. In this paper, Principal Component Analysis is performed on a number of environmental variables to create a scientifically weighted index that can be incorporated into composite welfare measures. Inclusion of the new environment index into the HDI greatly reduces the correlation between income and the index, as it incorporates environmental issues that are not significantly correlated with level of income.

Table of Contents

1. Introduction

Although traditional development research builds on economic variables such as Gross Domestic Product (GDP) per capita, it is today common to look at human development as a multidimensional concept, incorporating more than just economic factors. Well-being is now considered a much richer and more inclusive concept than purely measuring the level of income or economic growth rate. Considerable effort has been made to draw attention to other measures than the traditional monetary measures of national economic well-being achievement, and to better capture non-economic human welfare (McGillivray, 2005). Among a large diversity of indicators, measurements of social progress such as health and educational status are perhaps the most used in inter-country assessments of human development and are today available for some 160 countries (UNDP, 2003). The generally accepted Human Development Index (HDI) is based on only three indicators - income, health and education – to avoid making it too complex, and increase the chances for governments of poor countries to contribute with reliable data. It was created to emphasize that people and their capabilities, as opposed to economic growth alone, should be the ultimate criteria for assessing the development of a country (Sen, 1999). However, the method used for combining the variables into an index has received extensive critique for being subjective and lacking theoretical validity (see for example McGillivray, 1991).

While few would reject the intention to increase dimensionality in poverty reduction efforts and the assessment of well-being, these measures generally have one inherent flaw: they all tend to be highly correlated with pure income or production measures, and thereby not contribute much to the development debate. While the rankings of countries according to HDI to some extent differ from the ranking according to GDP per capita within the “extreme” groups: the most and the least developed countries, the average picture is the same for the two measures (McGillivray 1991). Using measures of GDP per capita and different welfare indicators for a sample of 173 countries, McGillivray (2005) finds that the rank order correlation between income and HDI amounts to 0.938, and -0.829 between income and the Human Poverty Index (HPI-1)[1].

[1] The Human Poverty Index is a composite index measuring dimensions of *deprivations* in quality of life (using variables measuring life expectancy, lack of basic education and lack of access to public and private resources) to estimate the extent of poverty in a community. The measure is derived separately for developing countries, HPI-1, and a group of high-income OECD countries, HPI-2. (UNDP, 1997)

Possibly related to the correlation of welfare measures with income, is the fact that many issues considered central to human development are not represented in the mentioned measures. One supposedly important aspect of human well-being is the environment and the sustainability of its use. In the field of environmental economics, increased human well-being is used as a benefit factor in estimating how much pollution abatement is desirable, given the costs of reductions (Welsch 2006). In 2010, the UNDP's Human Development Report Office (HDRO) together with researchers and representatives from international organizations held a meeting where the sustainability of the HDI was discussed. While recognizing that all parts of human development should be sustainable, special focus was put on the environment, since "the nature and the significance of the threats that we are facing are valid reasons to focus **primarly** [*sic!*] on **environment**." (UNDP-HDRO, 2010, p. 1) This is motivated by inter-generational considerations, as natural capital is not substitutable with other forms of capital to the same degree as man-made assets (UNDP-HDRO, 2010).

Indices of environmental sustainability and governance have become increasingly common, but no single one has yet received as much attention or use as the common welfare measures. These indices are usually focused solely on comparing environmental governance, and although some (for example the Environmental Sustainability Index, ESI) recognize that environmental sustainability is part of the Millennium Development Goals, they are rarely designed to be used together with development measures or on a year-to-year basis.

The aim of this paper is to create an index of environmental health and sustainability, that might be invoked in common welfare measures. *Is it possible to combine a number of environmental variables that can give a new dimension to the popular human development measures, and increase the ranking power of these beyond the regular income measures?*

Principal Component Analysis will be used for a more statistically relevant than subjective approach in creating an environmental index, from which a number of countries will be ranked according to reported values of each variable. Rankings based on a combination of the HDI and the environmental index will then be calculated. Rank correlations will be used to evaluate the correlations between the different indices and GDP per capita, to investigate whether incorporating an environment indicator decreases the high correlation of primarily the HDI and income per capita.

2. Performance Indices

2.1 The Human Development Index

The Human Development Index (HDI) was launched in the first Human Development Report, presented by the United Nations Development Program in 1990 (HDR 1990, UNDP).

The aim of the HDI is to measure three components of human well-being, or development:

- Long and healthy life, as measured by life expectancy
- Knowledge, as measured by mean years of schooling for adults aged 25, and expected years of schooling for children of school age
- A decent standard of living, as measured by GDP/capita

The index is built up as the geometric mean of the three components:

$$HDI = I_{\text{Health}}^{1/3} \times I_{\text{Education}}^{1/3} \times I_{\text{Living standards}}^{1/3} \qquad \text{(Equation 1)}$$

where each sub-index is created by the calculation:

$$I_{\text{component}} = \frac{\text{actual value} - \text{minimum value}}{\text{maximum value} - \text{minimum value}} \qquad \text{(Equation 2)}$$

The maximum value is the actual maximum observed for all countries over time (1980 to present). The minimum values are chosen as a “natural zero” or minimum subsistence value by the creators of the index, and can change from year to year. Minimum subsistence is defined as the minimum level “that a society needs to survive over time”. The minimum value for each variable is (for the index of 2010) set to:

- 0 years for both education variables
- 20 years for life expectancy
- 163 $ for GDP/capita

The index thereby makes the assumption that life expectancy of 20 years is sufficient to sustain a society, as well as a yearly income of 163 US dollars. This “sustainable income” measure is based on the lowest attained income per person ever: the record for Zimbabwe in 2008, equivalent to less than 45 cents per day (HDI Technical notes 2010).

The calculation procedure with absolute highs and "natural zeroes" gives an index that ranges from 0 to 1, with higher human development associated with a higher index score. In 2010, Norway and Australia scored highest on the HDI, with 0.938 and 0.937 respectively. The Democratic Republic of Congo and Zimbabwe are found at the bottom of the list, with 0.239 and 0.140 respectively.

Kovacevic (2010) gives an extensive review of the most common criticism expressed towards the HDI including a lack of dimensionality and oversimplifying the concept of human development and well-being by using too few or the wrong variables. Meanwhile, the creators of the index motivate the restriction on number of included variables with the need too keep costs of data collection for governments of poorer countries down. (Esty et al., 2005)

Of more technical and statistical consideration, is the notion that the weighting scheme in creating the index is arbitrary, and the fact that the index fails to capture information, as it is intrinsically correlated with pure income measures such as GDP/capita (Kovacevic, 2010). These correlations are high for absolute levels of income and human development indicators, as well as the rate of change over time for income and indicators (Ravallion 1997). McGillivray (1991) finds that the variables included in the HDI are highly correlated among themselves as well as with GDP/capita, and concludes that "…the index largely provides us with little more information regarding inter-country development levels than the more traditional indicator – GDP per capita, alone provides." McGillivray (1991) and Srinivasan (1994) both argue that the high inter-correlations pose a statistical problem, as the components of the HDI should be uncorrelated with each other to make the index contribute with as much information as possible.

Kovacevic (2010) performs a correlation analysis on the different components of the HDI, and find that the three components individually have correlations of between 0.71 (GDP) and 0.92 (Life expectancy) with the index. The logarithmic PPP[2]-adjusted GDP value (which is the form used in the actual creation of the index) has a correlation coefficient with the index of 0.94.

Kovacevic (2010) also performs PCA on the components of the HDI, and finds that the first PC accounts for 83 percent of total variation. The second PC has an eigenvalue lower than

[2] Purchasing Power Parity – used to adjust currency conversions to the local price level, by calculating what a unit of the currency can buy on the market. (See for example the OECD Statistics Directorate for definition).

one, and the author draws the conclusion that the HDI is one-dimensional, contrary to the beliefs of the creators of the index who describes it as a way of measuring several dimensions of poverty.

Another apparent drawback with the HDI is the design that keeps all scores between zero and one. This requires moving goalposts, as the maximum attainable value for an indicator must be increased every time a country beats the old record. Thereby, index scores are only comparable between those years when there was no change in maximum value for any indicator. It also means that a country that has improved its human development might actually score lower than the year before, if the goalposts have improved even more. Scores are therefore only comparable between countries, for the same year.

Although the HDI and other measurements of human welfare were not designed to determine international relations, development aid or other policy decisions, they are often used for prizing purposes[3] as well as in the allocation of aid on governmental level[4], and within the development-political debates of nongovernmental organizations (HDR, 1990 to 2006). It has also been proposed that a country's responsibility to take pollution reduction measures should be linked to its Human Development Index, in climate accords designed to follow the expiring Kyoto Agreement. (Wolff, 2011)

2.2 The Environmental Sustainability Index

In order to measure the sustainability in the management of each nation's environment and resources, the Yale Center for Environmental Law and Policy (YCELP) and the Center for International Earth Science Information Network (CIESIN) of Columbia University, in collaboration with the World Economic Forum and the Joint Research Centre of the European Commission developed the Environmental Sustainability Index (ESI). The index reportedly "benchmarks the ability of nations to protect the environment over the next several decades". The ESI is made up of five broad categories:

- Environmental Systems
- Reducing Environmental Stresses

[3] Since 2001 the pharmaceutical company Merck sells drugs at different prices with up to 90% discounts for countries that are classified into the 'low human development' category of the HDI, and 75% reductions for the 'medium human development' countries (Petersen and Rother, 2001 cited in Wolff, 2011).
[4] 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 cited in Wolff, 2011)

- Reducing Human Vulnerability to Environmental Stresses
- Societal and Institutional Capacity to Respond to Environmental Challenges
- Global Stewardship

which are comprised by 21 equally weighted indicators of environmental sustainability, which in turn are composed by 76 variables. These variables have been chosen after a thorough review of literature and available data, and consultation processes with policy-makers and experts, and include economic and quality of government indicators.

The designers of the index recognize big gaps in the data used, and note that "…any measure of sustainability will have shortcomings given the significant gaps in critical data sets…", but still consider the ESI a "useful gauge of national environmental stewardship". The authors also conclude that data unavailability is especially pronounced for developing countries, and remark that this must be addressed considering the environmental aims of the Millennium Development Goals.

The equal weighting of all 21 indicators is motivated by an aim for transparency, and that experts from the governmental, non-governmental and business sectors did not provide a united picture of which factors are most important. (Esty et al., 2005)

3. Creating an Index

3.1 Data Selection

The first problem that one faces in creating an index is the selection of relevant variables. This procedure is usually based on theoretical considerations, where the included data in combination is supposed to represent an aggregate measure, such as overall price level or economic development. When creating indices for international comparison on country-level, one important determinant in the choosing of variables is data availability. Reliable environmental, economic and political data is in general more accessible for economically developed countries and democracies (Jha and Murthy, 2006). Depending on which weighting scheme is used, the correlation of variables might also have to be taken into account, a question we will return to in the next section.

In choosing variables, one must also take into account the possibility of seasonality in data: the more often a variable is measured, the more likely are the measurements to capture any kind of seasonality (Vyas and Kumaranayake, 2006) such as an increase of expenditure on food during the months when the farmer's fields are less productive due to weather circumstances.

3.2 Weighting

Approaches used for assigning weights to the components of an index include equal weighting for each variable, weights arbitrarily chosen at researchers'/policy makers' discretion, and weighting according to the quality of data for each variable (see Mom Njong and Ningaye 2008 for a further discussion of weighting in poverty measurements). While these procedures allow the creators more freedom to choose and design their own index, the result can be of poor scientific and statistical quality. One severe drawback of discretionary assignment of weights is the risk of correlations and causality between variables in the index. If two variables are highly correlated, the probability that there is also a causal relationship increases, as one might expect there to be some mutual factor that drives both variables. Combining these variables in an index then increases the risk of giving too much weight to the one underlying factor that might drive both the measured variables (Jha and Murthy, 2004). If the intention is to give equal/arbitrary weights to each component in the index, one must therefore be careful in data selection not to include variables that are too highly correlated or show causality.

Principal component analysis (PCA) is a commonly used instrument for creating indices, as its main purpose is to reduce the dimensionality in data without losing relevant information. It is used to obtain coefficients that assign correct weights according to the statistical importance of each included variable in the index, and is increasingly used in welfare measurements. The next section presents the theoretical underpinnings of PCA, followed by PCA in practice and the use of principal components in creating indices.

4. Principal Component Analysis

4.1 Basic Concept

Analyzing multivariate structures requires techniques for reducing the many dimensions of a data set while keeping as much information as possible. Principal component analysis is one often used method to obtain a new set of variables that contain maximum amount of variation in the underlying multivariate data set.

The method of principal component analysis was first described during the first decades of the 20^{th} century (Jolliffe 2002), for the purpose of explaining a maximum amount of total variation in a set of underlying variables through "components" created by using correlation matrices. Even if much progress has been made since, the basic principles are still the same. The data used for performing principal component analysis is typically multivariate, with three or more interrelated variables. Covariance matrices are used to obtain a number of principal components expressed as linear relationships between the original variables. The first few principal components are supposed to retain the majority of variation in the original data.

Since the method of creating principal components is used for data sets with several variables, the data is often represented as a vector x of n random variables. To investigate the variation in these variables and what they have in common, the covariance matrix is of interest. (Jolliffe 2002)

The first step in performing principal component analysis is to create a maximum variance linear function of the n variables in the matrix x, on the form:

$$\xi_1 = \alpha_{11}x_1 + \alpha_{12}x_2 + \ldots + \alpha_{1n}x_n \qquad \text{(Equation 3)}$$

Where α is the weight of each variable x for the principal component ξ_1. The maximization of the variance is made subject to the restriction that the weights sum to 1. This condition is imposed to avoid increasing the variance just by ever increasing the value of the weights, and makes the solution independent of scale.

Next a second maximum variance linear function which is uncorrelated with the first is created:

$$\xi_2 = \alpha_{21}x_1 + \alpha_{22}x_2 + \ldots + \alpha_{2n}x_n \qquad \text{(Equation 4)}$$

This procedure is then repeated to create a system of maximum variance uncorrelated linear functions up to

$$\xi_n = \alpha_{n1}x_1 + \alpha_{n2}x_2 + \ldots + \alpha_{nn}x_n \qquad \text{(Equation 5)}$$

Thereby, each new principal component contains as much information as possible that is not accounted for by previous components, and the number of principal components obtained is the same as the number of variables in the data set.

The method of creating principal components as linear functions of the variables in the data set is equivalent to representing the original variables with respect to new axes in a coordinate system. Graphically, this is done by rotating the original axes to a new base which is orthonormal (and thus has a 90 degree angle between the new axes).

Algebraically, a point in the old coordinate system can now be expressed with respect to the new axes. A given point *y* with coordinates (x_1, x_2) in a system where vectors X_1* and X_2* are introduced as new base lines with angle θ to the original axes, X_1 and X_2 will receive the new coordinates

$$y = (\cos\theta\, x_1 + \sin\theta\, x_2)x_1^* + (-\sin\theta\, x_1 + \cos\theta\, x_2)x_2^* \qquad \text{(Equation 6)}$$

(Sharma, 1996)

4.2 Computations

The practical procedure of deriving PCs is based on computations of the covariance matrix of the original data. Using the vector X with n random variables, the covariance matrix is given by E(XX’) and denoted Σ. The *ij*th element of Σ is thus the covariance between variables *i* and *j*. Defining α' as a vector of weights for forming linear combinations of the original variables, gives the *k*th PC by

$$\xi_k = \alpha_k' x \qquad \text{(Equation 7)}$$

The variance of the new variable E(ξξ') is equal to $\alpha'\Sigma\alpha$, and is to be maximized subject to the constraint that $\alpha'\alpha = 1$.

One common way of maximizing the value of a function subject to a constraint is by using Lagrange multipliers, which is also the standard method for deriving principal components. This is done by expressing a function to be maximized:

$$\xi = \alpha'\Sigma\alpha - \lambda(\alpha'\alpha - 1) \qquad \text{(Equation 8)}$$

where λ is the Lagrange multiplier. Taking the first derivative of the above expression with respect to α gives a vector of partial derivatives defined as:

$$\frac{\partial\xi}{\partial\alpha} = 2\Sigma\alpha - 2\lambda\alpha \qquad \text{(Equation 9)}$$

Setting the derivative to zero yields:

$$\left(\Sigma - \lambda I_p\right)\alpha = 0 \qquad \text{(Equation 10)}$$

where I_p is the p×p identity matrix. λ is therefore an eigenvalue of the correlation matrix Σ, and α is the corresponding eigenvector (Jolliffe, 2002). (For a brief explanation of eigenvectors and eigenvalues in matrix algebra, see Appendix 1).

To decide which of the n eigenvectors give maximum variance, note that:

$$\alpha'\Sigma\alpha = a'\lambda a = \lambda a'a = \lambda \qquad \text{(Equation 11)}$$

Since $\mathrm{Var}(\alpha' x) = \alpha'\Sigma\alpha = \lambda$ the maximum variance is λ, the largest eigenvalue of the matrix, and α is the corresponding eigenvector.

5. PCA in Practice

5.1 Data Transformations

In creating principal components, correlation matrices are often used instead of covariances. This allows for comparison between PCAs of different data sets, as the result is independent of scale. It also makes the different variables within a data set independent of measurement units. PCA with covariance matrices gives the greatest weight to the variable with the numerically highest variance, which is only appropriate if the variables are measured using the same scale. Using covariances, the first PCs will be dominated by the variable(s) with the greatest variation due to differences in measuring units or high absolute values (which tend to give higher variance). The standardization of correlation matrices allows for PCs created from data sets with variables measured in different units, such as length, currencies and ordinal scales. The use of correlation matrices does not change the calculations in the above section (Jolliffe, 2002).

When standardizing data, outliers and skewness should first be taken care of as this will otherwise have distortionary effects in the standardizing process. Transformations will also solve the problem with different units for different variables (Kamanou, 2002). As variables in data samples are rarely normally distributed, several transformation options are available when non-normality is observed:

- Square root transformations
- Logarithmic transformations
- Inverse transformations

Square root transformation is performed by taking the square root of all values (it is often useful to also consider the fourth root, eighth root and other versions of the root transformation). The square root transformation requires non-negative values and preferably values above 1.00. The latter due to the fact that when applying the transformation, numbers between 0 and 0.99 behave differently than numbers of 1.00 and above (the square root of 4 is 2, but the square root of 0.4 is 0.63) and thus, you are treating some numbers differently than others which, in most cases, is not desirable (Osborne, 2002).

Logarithmic transformation can take on a number of bases. Baxter (1995) points out how many elements have a natural log-normal distribution, and natural or base ten logarithms are often proper transformations. Logarithmic transformations are all undefined for values below 1.00 and such variables are transformed by adding a constant and thus moving the minimum value of the distribution above the desired value.

Both square root and logarithmic transformations work by compressing the right side of the distribution more than the left side, and are thus effective on positively skewed distributions. When dealing with a negatively skewed distribution, it must first be reflected by multiplying the variables by -1.00 and then move them back to a minimum value of 1.00 by adding a constant before the transformation is applied. It must then be reflected again to restore the original order of the variables.

Inverse transformation of a variable Y means computing 1/Y. The inverse transformation reverses the order of the observations and, as in the case of negatively skewed distributions, one must be careful to reflect the distribution prior to applying the transformation so that the ordering after the transformation is identical to the original data. (Osborne, 2002)

When having a large number of observations all transformations are considered, comparing their effects on normality before choosing the transformation best for each variable. Table 1 (below) shows the mentioned transformations.

Square root transformations	***Logarithmic transformations***	***Inverse transformations***
$X^{0.5} = \sqrt{X}$	ln(X)	1/X
$X^{0.25} = \sqrt[4]{X}$	log10(X)	$1/X^{0.5} = 1/\sqrt{X}$
$X^{0.125} = \sqrt[8]{X}$		$1/X^{0.25} = 1/\sqrt[4]{X}$
		$1/X^{0.125} = 1/\sqrt[8]{X}$

Table 1. Some of the available data transformations

When normality is fulfilled, the data can be standardized and used in the correlation matrix to perform the PCA.

5.2 Using the Principal Components

The PCs can be interpreted as the dependent variable in a regular linear regression, and the weights on each variable as the coefficients for the independent variables. The difference is that in the use of Principal Components a measurable dependent variable is not available. Instead, the PC is created to account for as much unique information as possible in the observed variables. The first principal component will – by definition – account for the largest share of the total variance in the data, as principal components are created by maximizing the variance of the new variable (PC). Total amount of variation accounted for will be obtained by the cumulative percentage of variation accounted for by each added PC. As the main purpose of PCA is reducing dimensionality in the data to something more easily handled, one usually hopes that the data is correlated enough to produce a few PCs that account for much of the total variation in the data set. The more highly correlated the data is, the fewer PCs will be needed to explain much of the variation. (Jolliffe, 2002) The number of PCs to be used is mostly up to the discretion of the person performing the analysis and the purpose of the PCA, but using enough PCs to cover around 80 percent of total variance is a common rule of thumb. (Sharma, 1996) It is also possible to reduce the number of PCs using rules based on hypothesis testing, but such attempts are, according to Jolliffe (2002), often built on unrealistic assumptions about probability distributions and seem to retain more variables than necessary.

The use of PCA is greatly increased if the first few obtained PCs are easily interpreted, and thus gives an indication on what attribute each PC represents. The interpretation will depend on the ability to extract patterns in the weights given to the different variables in the data set. Jolliffe (2002) uses as an example data on anatomical measurements for a species, where the first PC usually has positive weights for all variables, and thereby represents some overall "size" of an individual of the species. In interpreting the PC coefficients, one must have in mind that they represent the variation among each other. Two variables with different signs will therefore represent a contrast. Given two variables A and B, where the coefficient of A is positive and the coefficient of B is negative in the same PC, one can draw the conclusion that some of the variation in the population comes from individuals with a high score on variable A, and a low score on variable B, or conversely a low score on A and a high score on B. Thus, the contrast between these two variables for each individual will be a source of variation in the population.

The numerical output from running PCA is usually defined to several decimals, which can give an expression of very exact results. However, when evaluating PCs the overall pattern should be concentrated on, as the interpretation of an exact numerical value of a certain weight is usually not very meaningful. Jolliffe (2002) generalizes the interpretation to just giving each variable a positive or negative sign, and only if the weight reaches some pre-determined cut-off value.

5.3 PCA as a Tool for Creating Indices

The problem with data availability for poor countries can be partly reduced by the use of PCA: since it reduces dimensionality in data to comprise a large number of variables into a smaller data set without loosing too much information, PCA can reduce the cost of data collection and thereby decrease the risk of low-quality data obtained from poorer countries, or states more likely to cheat in the data collection procedure (Jha and Murthy, 2006). PCA by definition explains how to preserve the overall information given in the original data, retaining a minimum number of variables.

The theoretical correlations of variables included in one index are instead of being a potential problem (due to causalities) an asset when using PCA. Since the procedure is based on correlations between variables, the index will for each variable sort out individual information that is not contained by the other variables.

The method of weighting the variables is automatically incorporated into the decision of using PCA: since the first principal component is usually thought to symbolize the overall index, the eigenvector obtained for the principal component is used as the weight.

5.4 Ranking Based on the First PC

As mentioned above, the first PC is usually interpreted as measuring overall “size” or incidence of the phenomenon of interest. By definition, the variables with higher variation between individuals (or in this case, countries) will be given more weight. On the other extreme, a variable that is equal for all countries will not be given any weight in the first PC, as this does not contribute to the variation in the population. Therefore, the weights from the first PC are often used for creating the index (See for example Vyas & Kumaranyake, 2006). McKenzie (2003) cites several studies suggesting that the first PC best discriminates between households in measuring welfare, and we have no reasons to believe that this is not the case also for countries in measuring environmental variables. Mom Njong and Ningaye (2008) and McGillivray (2005) also use the approach of letting the first PC set the weights for different welfare indices, even in cases with low eigenvalues for the first PC. Since an index is used for comparison between individuals or over time, discrimination or ranking is the main purpose: international comparison between countries is made possible by calculating a “score” from the observed individual values and corresponding weights for each variable. The exact value of this score is though not of great use: as mentioned above, the weight given to each variable in a PC is not readily interpretable, neither are the index scores obtained using them. The scores should therefore mainly be interpreted as a criterion for the ranking and distinguishing between countries, rather than as an exact measure (of environmental health or sustainability). Higher weight is given to variables that explain more variation and therefore are better at differentiating between countries, and the obtained index therefore is particularly good at ranking. The higher the eigenvalue for the first PC the more of total variation is retained in it, and thereby in the index.

Since the purpose of creating an environmental index is to incorporate some sustainability factor into the usual welfare measures, and evaluate correlations with the simple income measures, a proper method for comparing the new index to welfare indices in use and GDP per capita must be found. A method for computing rank order correlation is presented in the following section.

6. Evaluating the Index

6.1 Spearman's Rank Correlation Coefficient

When evaluating ranks or ranking procedures, Spearman's rank correlation coefficient is often used. The correlation coefficient, r_s, is defined as the correlation between two different rankings for one individual or observation when ranked twice, according to two different criteria, and is calculated by

$$r_s = 1 - \frac{6 \sum d_i^2}{n\,(n^2-1)} \qquad \text{(Equation 12)}$$

where d_i is the difference between rankings for each individual in the sample of n observations. A correlation of 1 implies perfect correlation between rankings: the individual receives the same rank according to both criteria. (Körner 1983)

A hypothesis test is formulated with

$$H_0 : \rho_s = 0 \qquad \text{(Equation 13)}$$

against a single or double-sided alternative hypothesis. The correlation coefficients are under the null hypothesis assumed to be symmetric around zero. For large *n*, critical values can be computed using a Student's t-test, on the form

$$t = \frac{r_s}{\sqrt{(1 - r_s^2)/(n-2)}} \qquad \text{(Equation 14)}$$

with $n - 2$ degrees of freedom. The accuracy of this approximation is increasing with the size of *n*, and can be used without need for corrections when $n \geq 100$ (Zar, 1972).

7. Data and Computations

The following section will present the data and sources, applied transformations and the computations for the combining of the indices.

7.1 Variables

The variables included in the index are chosen primarily according to two criteria: accessibility and relevance for the sustainability of a nation's and its population's welfare. For all variables, data is readily available through the UN database for the Millennium Development Indicators and the World Bank statistics database. There are few gaps in the data, and these are not especially prevalent for a specific group of countries (such as the least developed). Out of available data, variables that are usually present in other environmental indices and usually of concern in the sustainability debate, and represent different aspects of environmental degradation or protection are chosen. These can be divided into three main groups:

Atmospheric pollution

X2: Carbon dioxide (CO_2) emissions, metric tons per capita

X3: Consumption of ozone-depleting substances in ODP metric tons[5]

X4: Fossil fuel energy consumption, percent of total

X8: Methane (CH_4) emissions, thousand metric tons of CO_2 equivalent

X9: Nitrous oxide (N_2O) emissions, thousand metric tons of CO_2 equivalent

Land and water pollution

X5: Use of fertilizers per 1000 hectares of agricultural land area, nitrogen (metric tons)

X6: Use of fertilizers per 1000 hectares of agricultural land area, phosphate (metric tons)

X7: Use of fertilizers per 1000 hectares of agricultural land area, potash (metric tons)

X10: Sulfur dioxide (SO_2) emissions, thousand metric tons

Environmental protection efforts

X1: Proportion of land area covered by forest, percentage

X11: Proportion of terrestrial and marine areas protected, percentage

X12: GEF benefits for biodiversity[6]

[5] Ozone depleting substances measured in metric tons, each weighted by its ozone-depletion potential

Using dummy variables and categorical data greatly increases the complexity of the PCA procedure, although there are proposed methods for using non-continuous data. However, data on this form will not be used here, to keep the simplicity of the PCA procedure.

Data from 2005 for all variables is available for 114 countries. Variables that are not expressed in proportions are divided by the size of the country's population, as per capita measures allows for an analysis without taking the size of the country into account. The variable X3 – Consumption of Ozone-depleting substances – lack data for 25 European Union members, as this variable is only reported as an aggregate for the EU. To account for the missing observations in the index, the mean value of X3 is computed and per capita-adjusted for each EU member.

For comparison with income, GDP per capita in PPP-adjusted prices from 2005 is used since it is the method used in creating the HDI as well as in evaluations of welfare measures (such as McGillivray 2005, Kovacevic 2010).

7.2 Transformations

None of the 12 variables are normally distributed and therefore transformations of the data must be performed (see Appendix 2 for histograms of the variables). Variable X4 has a negative skewness while variables X1 - X3 and X5 - X12 all show positively skewed distributions to some extent. Many variables also show extreme outliers needed to be taken care of before standardizing. Applying all transformations from Table 1 (section 5.1) and comparing their effects on normality leads to the choice of transformations shown in Table 2 below (see Appendix 3 for histograms of the transformed variables).

---

[6] World Bank measure of biodiversity, "based on the species represented in each country, their threat status, and the diversity of habitat types in each country".

Variable	***Y=f(X)***	***Variable***	***Y=f(X)***
X1	$Y1 = \sqrt{X1}$	X7	$Y7 = \sqrt[8]{X7}$
X2	$Y2 = \sqrt[8]{X2}$	X8	$Y8 = \frac{1}{\sqrt[4]{X8}}$
X3	$Y3 = \sqrt[4]{X3}$	X9	$Y9 = \frac{1}{\sqrt[4]{X9}}$
X4	$Y4 = \sqrt[4]{X4}$	X10	$Y10 = \sqrt[8]{X10}$
X5	$Y5 = \sqrt[8]{X5}$	X11	$Y11 = \sqrt{X11}$
X6	$Y6 = \sqrt[8]{X6}$	X12	$Y12 = \sqrt[8]{X12}$

Table 2. Applied data transformations

Since variable X4 showed a negatively skewed distribution the variable was reflected by multiplying by -1.00 and added to a constant to bring values back above 1.00 prior to performing the transformation. It was then reflected back to original order by multiplying by -4.00 again. Since the inverse transformation chosen for variables X8 and X9 changes the order of the observations, it was also reflected by multiplying by -1.00 and added to a constant to bring the values back above 1.00.

The distribution of each variable and whether near-normality was achieved was assessed using the histograms in Appendix 3.

7.3 Estimations

The PCA is performed on the transformed variables, using the default standardization process for PCA in SAS 9.2 to account for different scales on different variables.

The resulting first principal component is then assessed to make sure that all weights make logical sense (have the right sign). One would expect the “degradation variables” (pollution, use of fertilizers) to show a negative sign, and the “protection variables” to have positive coefficients if the first principal component behaves according to theory and actually represents overall environmental sustainability. Variables with higher variation across countries should be given higher weights, as they are supposed to be better at differentiating.

The index scores for each country are then calculated by the formula

$$\text{Score} = \sum_{i=1}^{12} w_i y_i \qquad \text{(Equation 15)}$$

after which all countries are ranked according to their index scores.

7.4 Evaluation and Comparison

The rankings from the environmental index are compared to those of the HDI and PPP-adjusted GDP for 2005, using Spearman’s rank correlation coefficient. The environmental index and the HDI are then combined to perform a new ranking of countries based on both the classical development indicators and the environmental index. The combining of the indices are performed using two different methods. The first incorporates the environmental index as a fourth indicator in the HDI, and is thus given an “arbitrary” weight of 1/4:

$$I_{combined} = 0.25 I_{sustainability} + 0.75 I_{HDI} \qquad \text{(Equation 16)}$$

This requires that the environmental index is distributed between zero and one, for which the procedure of the HDI is followed:

$$I_{sustainability} = \frac{\text{actual value - minimum value}}{\text{maximum value - minimum value}} \qquad \text{(Equation 17)}$$

The second method performs PCA on the environmental index together with the four HDI variables (life expectancy, expected years of schooling, mean years of schooling and PPP-adjusted GDP) to obtain their respective weights when computing the combined index score.

The rankings from these two combinations of the environmental index and the HDI are then compared to that of the original HDI and PPP-adjusted GDP.

8. Results

The PCA on the environmental variables gives an eigenvalue of 4.58 to the first PC, which explains 38 percent of total variation.

	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1.	4.58191555	2.42186136	0.3818	0.3818
2.	2.16005419	0.58434427	0.1800	0.5618
3.	1.57570991	0.56425107	0.1313	0.6931
4.	1.01145885	0.41926889	0.0843	0.7774
5.	0.59218996	0.02775214	0.0493	0.8268
6.	0.56443782	0.03680833	0.0470	0.8738
7.	0.52762950	0.15241015	0.0440	0.9178
8.	0.37521935	0.12014450	0.0313	0.9491
9.	0.25507485	0.08215146	0.0213	0.9703

Table 3. Eigenvalues and proportion of variation explained by the obtained PC's

All weights have the opposite signs from expected – the PC is indicating a negative environmental trend. However, by just switching all signs, an index that give positive weights to amount of forest and bio-protection while having negative signs for all pollution variables is achieved.

	Prin1		**Prin1**
Y1	-.199573	Y7	0.359673
Y2	0.365331	Y8	0.029420
Y3	0.304604	Y9	0.124584
Y4	0.350387	Y10	0.263487
Y5	0.395497	Y11	-.195396
Y6	0.378990	Y12	-.240294

Table 4. Eigenvectors for the first principal component

The eigenvectors (with reversed signs) of the first principal component give the weights to be used when forming the index with each of the 12 variables, $Y_1 – Y_{12}$:

Variable	Weight
Proportion of land area covered by forest	0.199573
CO_2 emissions	-0.365332
Fossil fuel consumption, percent of total	-0.350387
Consumption of ozone-depleting substances	-0.304604
Use of nitrogen	-0.395497
Use of phosphate	-0.378990
Use of potash	-0.359673
Methane emissions	-0.029420
Nitrous oxide emissions	-0.124584
Sulfur dioxide emissions	-0.263487
Terrestrial and marine areas protected	0.195396
Biodiversity	0.240294

Table 5. Weights for each variable

$$\text{Index} = 0.199573Y1 - 0.365331Y2 - 0.304604Y3 - 0.350387Y4 - 395497Y5 - 0.378990Y6 - 0.359673Y7 - 0.029420Y8 - 0.029420Y9 - 0.263487Y10 + 0.195396Y11 + 0.240294Y12 \quad \text{(Equation 18)}$$

The computations give a score for each country ranging from 3.44983 for Zambia to -0.45824 for Qatar, where a higher index score is associated with higher environmental sustainability (see Appendix 4 for all countries' scores and their respective rankings). The exact value is, as mentioned in section 6, not to be seen as a measure of environmental health or sustainability, but rather just as a criterion for the ranking between countries.

The ranking according to the environmental index shows no significant correlation with either the original HDI or PPP-adjusted GDP per capita for 2005. The rank-order correlation between the HDI and PPP-adjusted GDP per capita is however very high, with a value of 0.95.

	Env. Index	**HDI UNDP**	**GDP (PPP)**
Env. Index	1.00000	-0.17191 0.0739	-0.10830 0.2601
HDI UNDP	-0.17191 0.0739	1.00000	0.94744 <.0001
GDP (PPP)	-0.10830 0.2601	0.94744 <.0001	1.00000

Table 6. Correlations: Environmental index, HDI and income

When transforming the environmental index and incorporating it into the HDI using the first method with the weighting of 1/4, new rankings with values ranging from 0.87084 to 0.28452 are achieved. Australia and New Zealand are now found at the top and Cameroon, Kenya and Senegal are found at the bottom (see Appendix 5 for all countries' scores and their respective rankings).[7] The rankings from this combined index show significant correlations of 0.72 and 0.63 with HDI and PPP-adjusted GDP respectively.

	Combined	**HDI UNDP**	**GDP (PPP)**
Combined	1.00000	0.71729 <.0001	0.63463 <.0001
HDI UNDP	0.71729 <.0001	1.00000	0.94744 <.0001
GDP (PPP)	0.63463 <.0001	0.94744 <.0001	1.00000

Table 7. Correlations: the ¼-weighted combined index, HDI and income

[7] Missing values are present for 30 countries due to missing values in the UNDP HDI scores. Rank-order correlations are computed taking the missing values into account . This is also the case when computing correlations between the environmental index and HDI (Table 6).

When using the second method for combining the two indices, the PCA for the environmental index and the HDI variables gives an eigenvalue of 2.97 to the first PC, which explains 59 percent of total variation.

	Eigenvalue	**Difference**	**Proportion**	**Cumulative**
1	2.96752387	2.04135749	0.5935	0.5935
2	0.92616638	0.33855836	0.1852	0.7787
3	0.58760802	0.22124241	0.1175	0.8963
4	0.36636561	0.21402948	0.0733	0.9695
5	0.15233612		0.0305	1.0000

Table 8. Eigenvalues and proportion of variation explained by the obtained PC's

The eigenvectors of the first principal component give the weights to be used when forming the combined index with each of the five variables (HDI1 = Life expectancy, HDI2 = Expected years of schooling, HDI3 = Mean years of schooling and HDI4 = the logarithmic PPP-adjusted GDP per capita).

	Prin1
HDI1	0.499323
HDI2	0.539759
HDI3	0.500167
HDI4	0.407380
Env. Index	0.395497

Table 9. Eigenvectors for the first principal component

$$I_{combined} = 0.499323 I_{HDI1} + 0.539759 I_{HDI2} + 0.500167 I_{HDI3} + 0.407380 I_{HDI4} + 0.395497 I_{sustainability} \quad \text{(Equation 19)}$$

Computing the scores yield values ranging from 61.9482 to 36.4265. The PCA-based combined index show similar rankings to the ¼- weighted index with Australia and New Zealand on top and Cameroon, Senegal and Kenya at the bottom. The PCA-based combined index also show significant rank-order correlations with HDI and PPP-adjusted GDP per capita although they are slightly lower (0.71 and 0.62 for HDI and PPP-adjusted GDP respectively).

	Combined	**HDI UNDP**	**GDP (PPP)**
Combined	1.00000	0.70530 <.0001	0.62462 <.0001
HDI UNDP	0.70530 <.0001	1.00000	0.94744 <.0001
GDP (PPP)	0.62462 <.0001	0.94744 <.0001	1.00000

Table 10. Correlations: the PCA-based combined index, HDI and income

Rank-order correlations between the two different combined indices equals 0.99 indicating that changes in rankings between the indices are very small and thus, that the two weighting methods in this case yield approximately the same result.

	¼-Weighted Combined	**PCA Combined**
¼-Weighted Combined	1.00000	0.99291 <.0001
PCA Combined	0.99291 <.0001	1.00000

Table 11. Correlations: the ¼-weighted combined index and the PCA-based combined index

9. Analysis and Discussion

The claim that HDI and (PPP-adjusted) income are highly correlated is supported in our analysis and adding an environmental index to the HDI decreases the correlation with GDP, but keeps it positive and significant. This is not considered a problem, as income is a very relevant estimator of welfare, if not the only one.

The non-significant correlations between the environment index and HDI and income could possibly mean that the new information in the index could bring a new dimension to common welfare measures such as the HDI. Since incorporating the environmental index in the HDI decreases the correlation, this indicates that countries with higher income receive a worse ranking according to the environmental index than according to the HDI. Related to this, it is interesting to note that scores are lower for the transformed 1/4-weighted combined index than for the HDI: the highest combined score is 0.871, while the HDI reaches 0.938. Since 1 is the highest (theoretically) attainable score for both indices, the lower combined scores is likely due to the added dimension which works in opposite direction from the other indicators in the index.

The use of PCA in the creation of the environmental index underscores how the use of arbitrary weights in creating indices might distort results due to correlations that give some indicators too much importance. Even if some 50 percent of the weights are of quite similar magnitude in the environmental index, in particular methane and nitrous oxide emissions have low weights, indicating that their respective contributions to variation in environmental health are small. In order to reduce data collection efforts, one or both of these variables might be left out without any significant reduction of explanatory value of the index. The use of principal components in creating the index assures that there is zero correlation between the variables included in the index. Equal weighting of all variables would give some variables more importance than motivated by their statistical contribution to an overall environment indicator.

Looking at the rankings from the environmental index, many low-or middle income countries are found at the top. These countries have a combination of fairly low emissions, combined with an abundance of natural resources such as forests and marine habitats. The countries at the bottom of the index are largely located in the middle-east and North Africa. These nations are generally oil-producers (and thereby largely oil-consumers) covered by deserts, two factors that contribute negatively to the index score. The rankings seem plausible considering these extremes, while the large group of countries in the middle are harder to distinguish according to specific variables.

The weighting structure to incorporate the environmental index into the HDI giving it 1/4 of the total score could be exposed to the same critique as the weighting scheme of the HDI. It is however a conscious choice, made to follow the construction of the HDI to allow for clear comparisons. The option of also performing the PCA on the index together with the HDI indicators gives more statistically defendable weights but as rank-order correlation shows, the both indices rank countries in a very similar manner. The transformations of index scores to a value between zero and one within the 1/4-weighted index must also be criticized for shifting goalposts from year to year and not allow for clear comparisons over time, but is unfortunately also a necessary measure to achieve a new index that is comparable to the HDI. The lack of time dimension must also be addressed, since the analysis only covers one year. It is however likely that the results presented here are consistent over time, as there is unlikely to be any seasonality present in yearly data and the PCA procedure does not take any absolute values into account. The results are also generalizable to other welfare measures, as the creation of the environmental index is not dependent on the structure of the HDI itself.

The general consensus that adding more variables to the HDI can decrease its rank correlations with income is supported by the analysis in this paper. This could possibly be done by choosing any more or less random set of variables to add and pre-check their correlations with income, but considered the increased importance of environmental factors in the development debate, an environmental index on the form presented here could be a possibility to consider when the dimensionality of the HDI is to be increased.

References

Baxter, M. J. (1995), *Standardization and Transformation in Principal Component Analysis, with Applications to Archaeometry*. Journal of the Royal Statistical Society: Series C (Applied Statistics), Vol. 44, Issue 4, pp. 513-527.

Cleveland, W. S. (1984), *Graphical methods for data presentation*: *Full scale breaks, dot charts, and multibased logging*. The American Statistician, Vol. 38, Issue 4, pp. 270- 280.

Esty, D. C., Levy, M., de Sherbinin, A., Srebotnjak, T. (2005), *2005 Environmental Sustainability Index*: *Benchmarking National Environmental Stewardship*. New Haven: Yale Center for Environmental Law & Policy.

Jerrold H. Z. (1972), *Significance Testing of the Spearman Rank Correlation Coefficient*. Journal of the American Statistical Association, Vol. 67, Issue 339, pp. 578-580.

Jha, R., Murthy, B. (2004), *A critique of the environmental sustainability index*. Australian National University, Arndt-Corden Department of Economics, Working paper series, 2003-08.

Joliffe, I. T. (2002), *Principal Component Analysis*. Second edition, New York: Springer-Verlag, New York, Inc.

Kamanou, G. (2002), *Combining development indicators using an Iterative-Principal Component Analysis*. United Nations Statistics Division, Joint statistical meetings - Section on Government Statistics. Retrieved April 4, 2011 from http://www.amstat.org/sections/srms/Proceedings/y2002/files/JSM2002-001154.pdf

Kovacevic, M. (2010), *Review of HDI Critiques and Potential Improvements*. Human Development Research Paper 2010/33, United Nations Development Programme.

McGillivray, M. (1991), *The Human Development Index*: *Yet Another Redundant Composite Development Indicator?* World Development, Vol. 19, Issue 10, pp. 1461–1468.

McGillivray, M. (2005), *Measuring Non-Economic Well-Being Achievement*. Review of Income and Wealth, Vol. 51, Issue 2, pp. 337-364.

McKenzie, D. J. (2005), *Measuring Inequality with Asset Indicators*. Journal of Population Economics, Vol. 18, Issue 2, pp. 229-260.

Mom Njong, A. and Ningaye, P. (2008), *Characterizing Weights in the Measurement of Multidimensional Poverty*: *An Application of Data-Driven Approaches to Cameroonian Data*. OPHI Working Papers ophiwp21, Queen Elizabeth House, University of Oxford.

Osborne, J. (2002), *Notes on the use of data transformations*. Practical Assessment, Research & Evaluation, Vol. 8, Issue 6. Retrieved May 8, 2011 from http://PAREonline.net/getvn.asp?v=8&n=6

Ravallion, M. (1997), *Good and Bad Growth*: *The Human Development Report*. World Development, Vol. 25, Issue 5, pp. 631-638.

Sen, A. (1999), *Development as Freedom*. Oxford: Oxford University Press.

Sharma, S. (1996), *Applied Multivariate Techniques*. John Wiley & Sons, Inc.

Srinivasan, T. N. (1994), *Human Development*: *A New Paradigm or Reinvention of the Wheel?* The American Economic Review, Vol. 84, Issue 4, pp. 238-243.

UNDP (1990), *Human Development Report*: *Concept and Measurement of human development.*
UNDP (1997), *Human Development Report*: *Human Development to Eradicate Poverty*.
UNDP (2003), *Annual Report 2003*: *A World of Development Experience*.
UNDP (2010), *Human Development Report 2010 Technical Notes*.
UNDP-HDRO (2010), Summary on *Sustainability, environment and human development consultation UNDP-HDRO*, London, 1st and 2nd Feb 2010.
Vyas, S., Kumaranayake, L. (2006), *Constructing socio-economic status indices*: *how to use principal components analysis*. Health Policy Plan, Vol. 21, Issue 6, pp. 459–468.
Welsch, H. (2007), *Environmental welfare analysis*: *A life satisfaction approach*. Ecological Economics, Vol. 62, Issue 3-4, pp. 544-551.
Wolff, H., Chong, H. and Auffhammer, M. (2011), *Classification, Detection and Consequences of Data Error*: *Evidence from the Human Development Index*. The Economic Journal (Early view), DOI: 10.1111/j.1468-0297.2010.02408.x

Databases

UN database for the Millennium Development Indicators, http://mdgs.un.org/unsd/mdg/
World Bank statistics database, http://data.worldbank.org/

Appendix 1

Eigenvectors and eigenvalues in matrix algebra

The concepts of eigenvectors and eigenvalues are important for the understanding of the derivation of principal components. In matrix algebra, an eigenvector is defined as a vector of n elements, which changes only scale – and thus not direction – when multiplied by a matrix.

The vector:

$$\mathbf{X} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

has a direction and a magnitude (scale), which is easily seen if plotted in a coordinate system. If, when multiplying the vector with a given matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

only the magnitude and not the direction of the vector changes, then the vector is an *eigenvector* of that particular matrix. The factor change of magnitude is the *eigenvalue* of the vector.

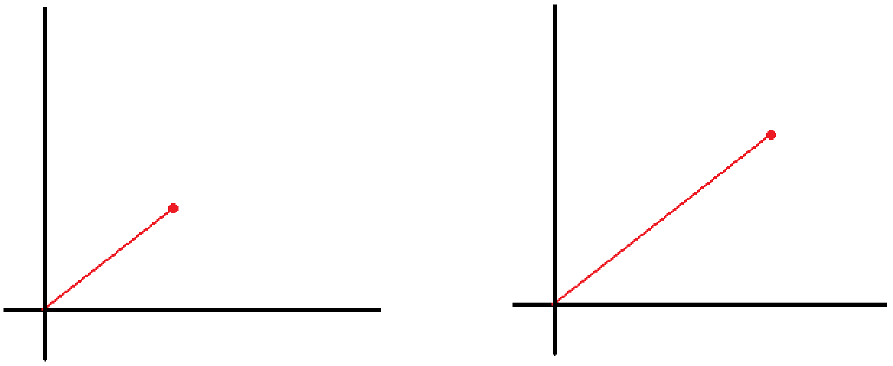
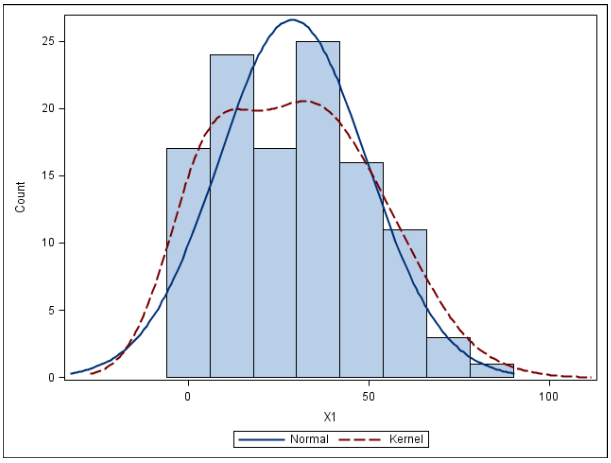
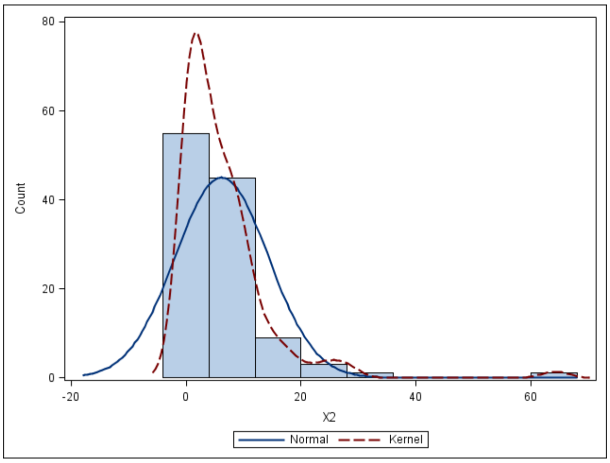


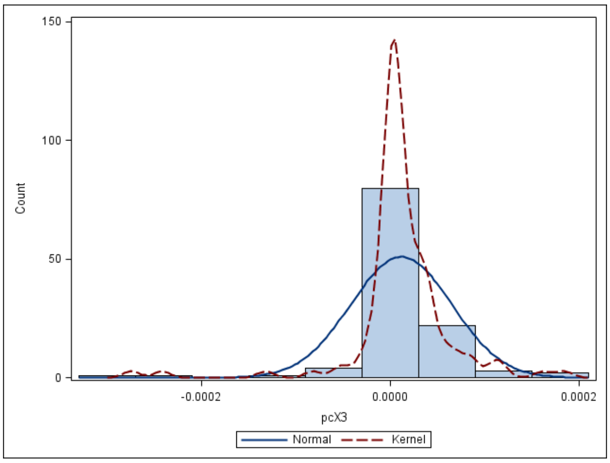
Figure 1. If multiplication of a vector with a matrix only results in the change of scale of the vector, it is an eigenvector of the particular matrix.

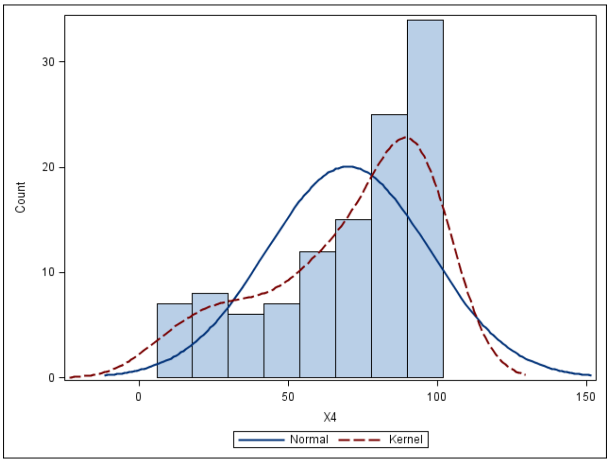
Appendix 2

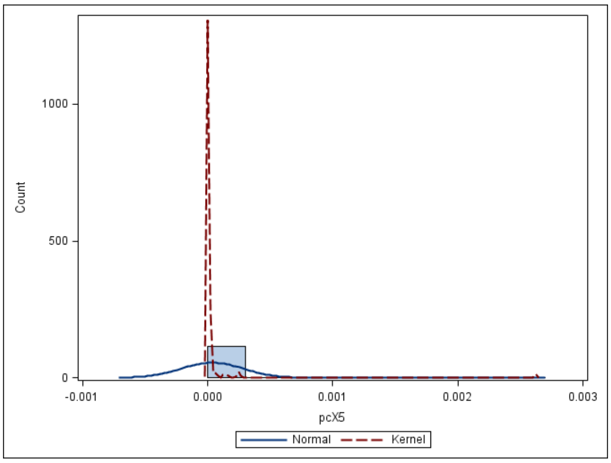
<u>Histograms of original data</u>

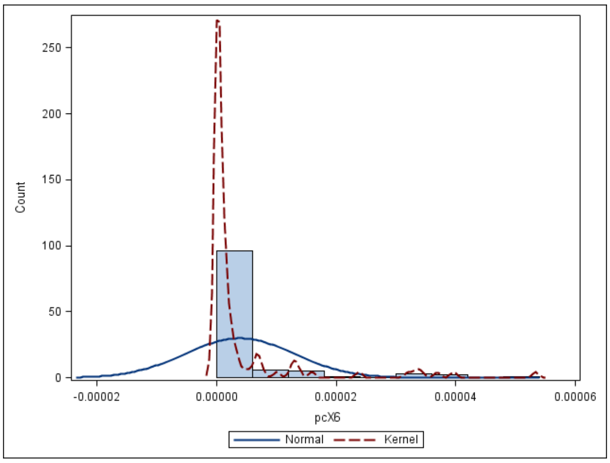


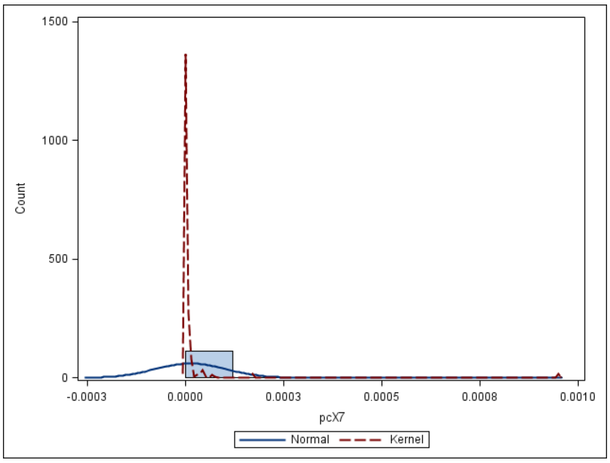




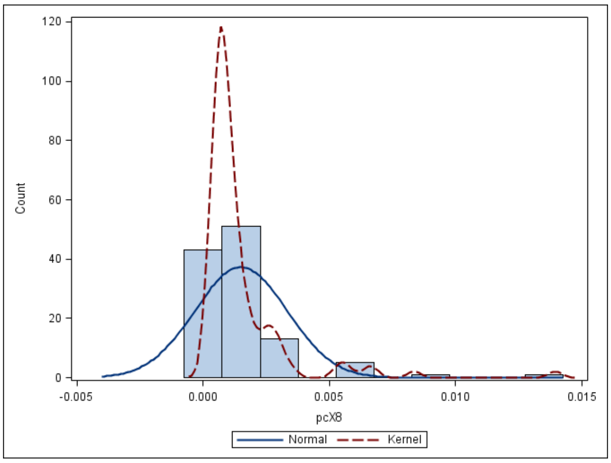




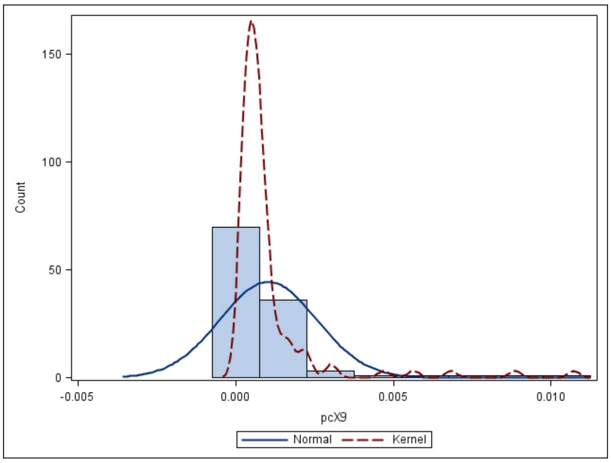




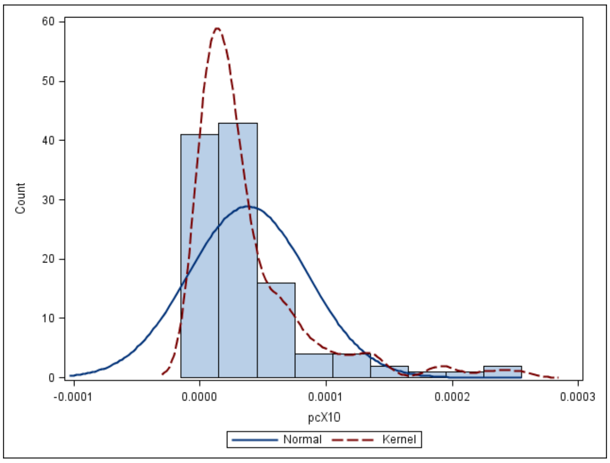
1500
1000
500
0
Count
-0.0003
0.0000
0.0003
0.0005
0.0008
0.0010
pcX7
Normal
Kernel



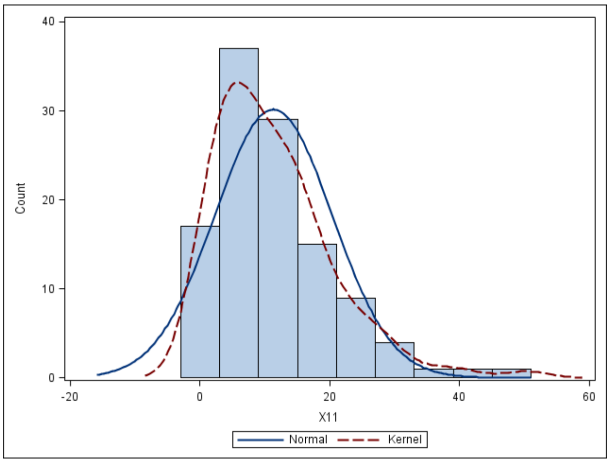
120
100
80
60
40
20
0
Count
-0.005
0.000
0.005
0.010
0.015
pcX8
Normal
Kernel



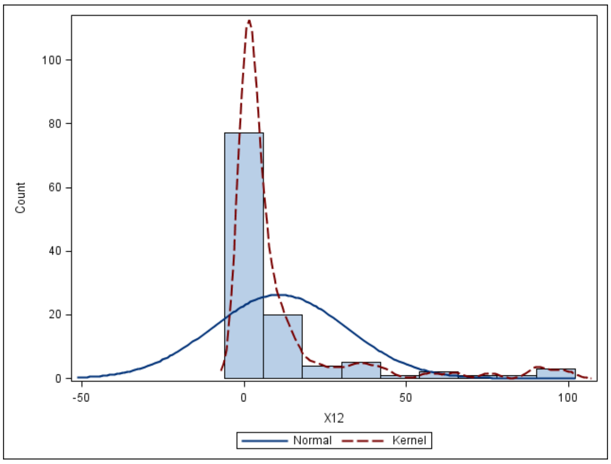
150
100
50
0
Count
-0.005
0.000
0.005
0.010
pcX9
Normal
Kernel



60
50
40
30
20
10
0
Count
-0.0001
0.0000
0.0001
0.0002
0.0003
pcX10
Normal
Kernel



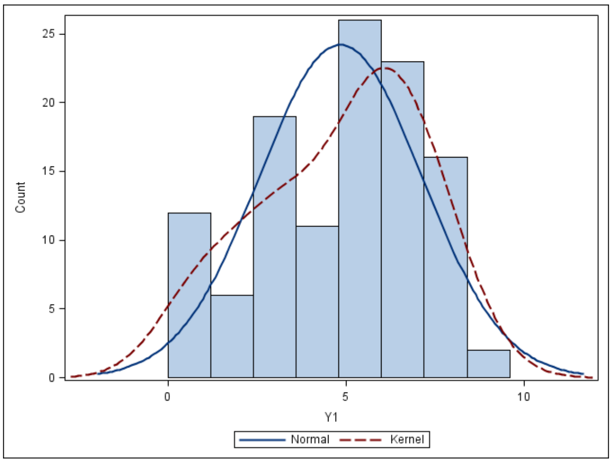
40
30
20
10
0
Count
-20
0
20
40
60
X11
Normal
Kernel

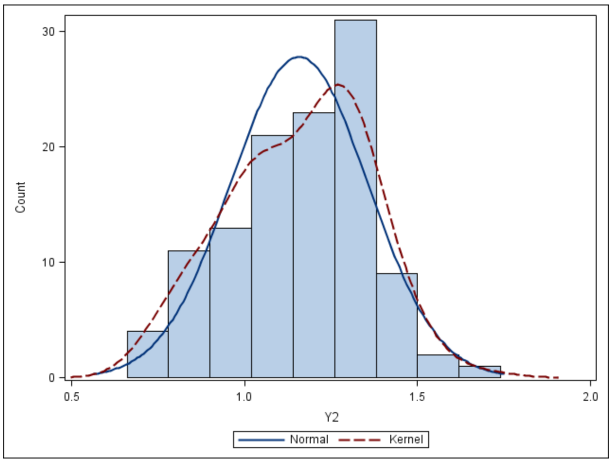


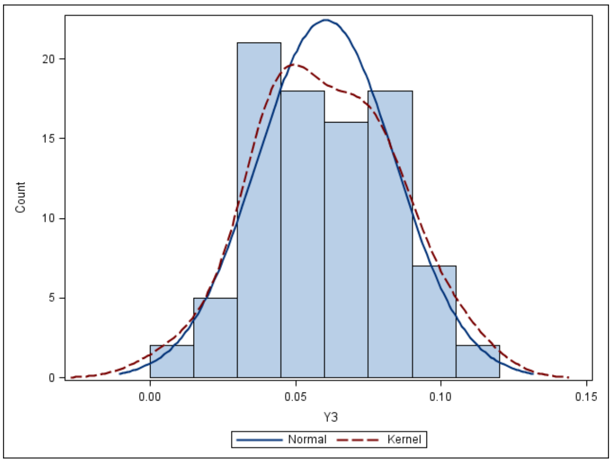
100
80
60
40
20
0
Count
-50
0
50
100
X12
Normal
Kernel

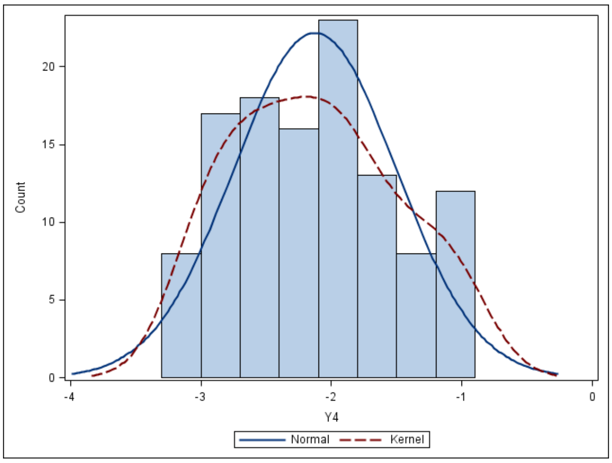
Appendix 3

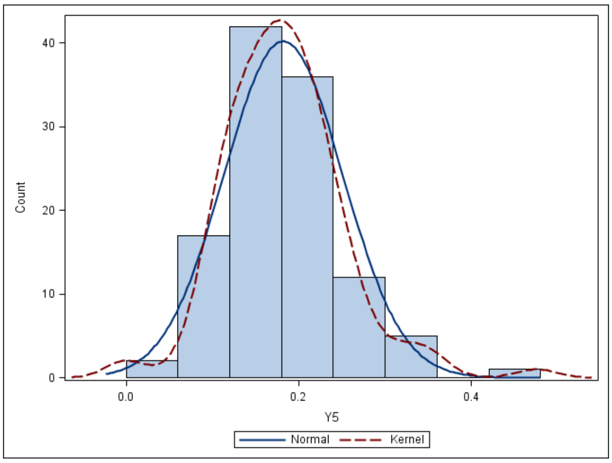
Histograms of transformed data

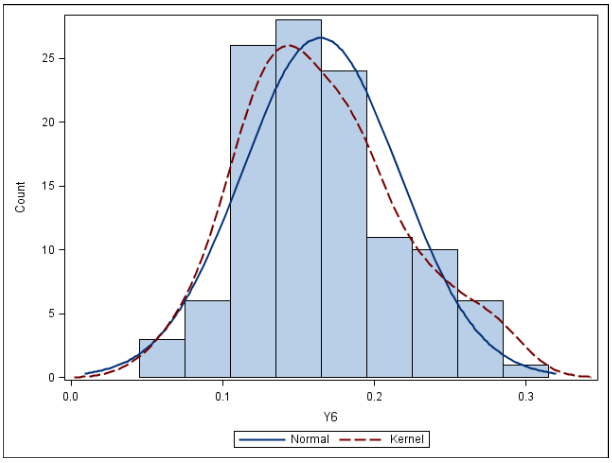


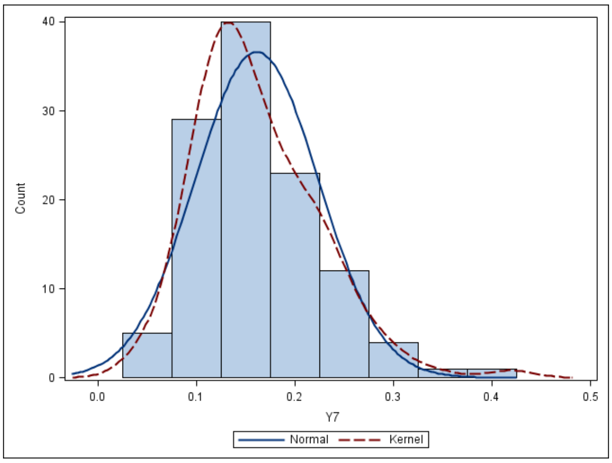


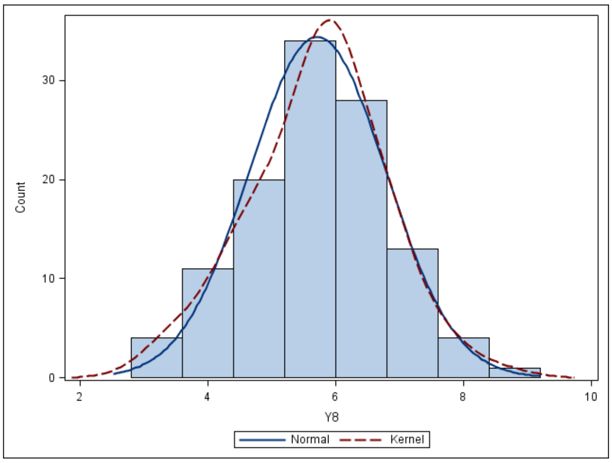


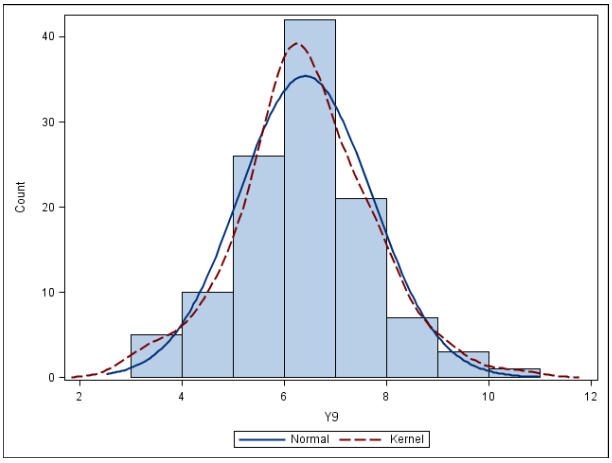


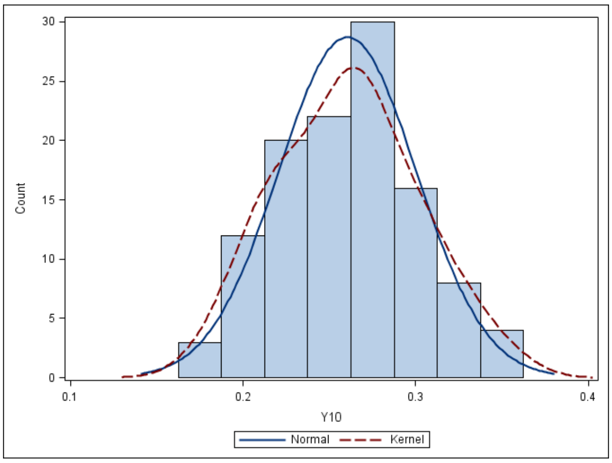


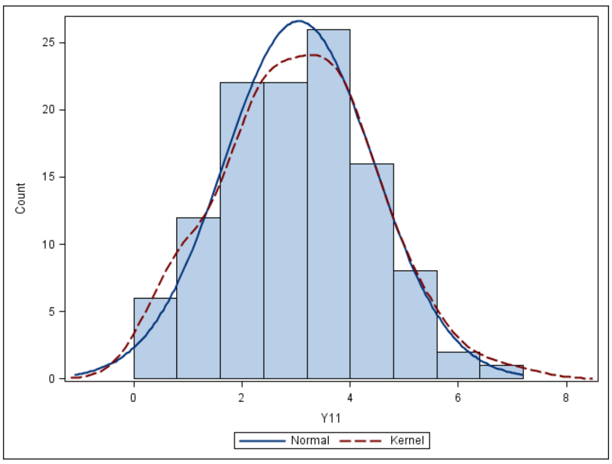


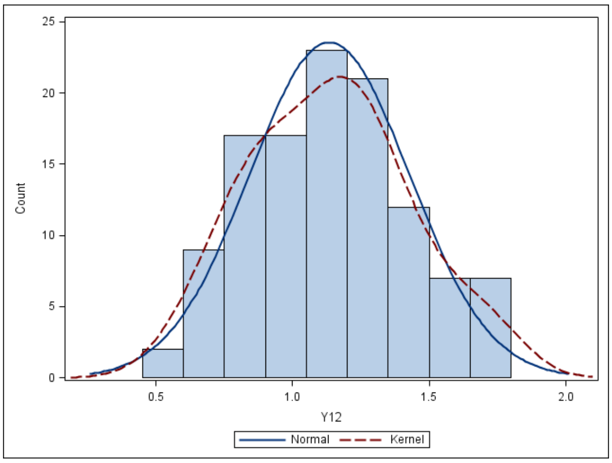

Count
Y7
Normal
Kernel


Count
Y8
Normal
Kernel


Count
Y9
Normal
Kernel


Count
Y10
Normal
Kernel


Count
Y11
Normal
Kernel


Count
Y12
Normal
Kernel

Appendix 4

Ranking according to the PCA-based environmental sustainability index

Rank	Country	Score	Rank	Country	Score
1	Zambia	3.44983	33	Latvia	2.38073
2	Cambodia	3.21636	34	Malaysia	2.37307
3	Brazil	3.17255	35	Slovenia	2.36794
4	Gabon	3.15337	36	Slovakia	2.34448
5	Tanzania	3.14330	37	Ecuador	2.32596
6	Venezuela	3.10077	38	Sudan	2.27923
7	Zimbabwe	2.99294	39	New Zealand	2.27126
8	Colombia	2.91243	40	Vietnam	2.24367
9	Congo, Rep.	2.87328	41	Estonia	2.24209
10	Guatemala	2.83468	42	Russia	2.21983
11	Senegal	2.81885	43	United States	2.19600
12	Angola	2.77208	44	France	2.19424
13	Cote d'Ivoire	2.76269	45	Mexico	2.15224
14	Cameroon	2.73549	46	Thailand	2.14203
15	Bolivia	2.73306	47	China	2.08960
16	Costa Rica	2.68690	48	South Korea	2.07589
17	Peru	2.67167	49	Georgia	2.07162
18	Mozambique	2.64722	50	Brunei Darussalam	2.02229
19	Sweden	2.63991	51	India	1.97204
20	Myanmar	2.59613	52	Italy	1.94535
21	Nepal	2.58385	53	Chile	1.94281
22	Panama	2.57932	54	Kenya	1.94184
23	Japan	2.57528	55	Spain	1.93573
24	Finland	2.51232	56	Philippines	1.93557
25	Honduras	2.49965	57	Czech Republic	1.93119
26	Germany	2.49220	58	Canada	1.92889
27	Indonesia	2.47534	59	Bulgaria	1.92725
28	Paraguay	2.43276	60	Norway	1.91374
29	Austria	2.42915	61	Poland	1.90326
30	Ghana	2.39945	62	Albania	1.88013
31	Sri Lanka	2.39507	63	Eritrea	1.87288
32	Ethiopia	2.38570	64	Togo	1.85382

65	Namibia	1.85188	90	Saudi Arabia	1.12979
66	Portugal	1.84227	91	Denmark	1.12758
67	Lithuania	1.82893	92	Israel	1.12095
68	Jamaica	1.79061	93	Uruguay	1.06158
69	Australia	1.74763	94	Cyprus	1.02757
70	Croatia	1.71959	95	Morocco	1.01589
71	Cuba	1.71767	96	Iran	1.01209
72	Luxembourg	1.70567	97	Iceland	0.99942
73	Belarus	1.66324	98	Moldova	0.92611
74	Greece	1.63936	99	Tunisia	0.91606
75	United Kingdom	1.58837	100	Lebanon	0.76429
76	Trinidad and Tobago	1.57847	101	Ireland	0.72162
77	Armenia	1.54958	102	Algeria	0.64662
78	Bosnia and Herzegovina	1.47806	103	Jordan	0.61812
79	Pakistan	1.46019	104	Egypt	0.59351
80	Hungary	1.45842	105	Kazakhstan	0.48216
81	El Salvador	1.44153	106	Yemen	0.46168
82	Argentina	1.42161	107	United Arab Emirates	0.40492
83	Mongolia	1.40746	108	Syrian Arab Republic	0.40158
84	Ukraine	1.33914	109	Oman	0.35923
85	Netherlands	1.26342	110	Singapore	0.14337
86	Bangladesh	1.24776	111	Malta	0.00919
87	Turkey	1.21496	112	Libya	-0.02069
88	Tajikistan	1.16894	113	Bahrain	-0.18306
89	Azerbaijan	1.16596	114	Qatar	-0.45824

Appendix 5

Ranking according to the 1/4-weighted and PCA-based combined indices

1/4-Weighted Combined Index			***PCA-Based Combined Index***		
Rank	**Country**	**Score**	**Rank**	**Country**	**Score**
1	Australia	0.87084	1	Australia	61.9482
2	New Zealand	0.86004	2	New Zealand	61.3267
3	Iceland	0.85926	3	Norway	60.1124
4	Norway	0.85891	4	Iceland	59.9394
5	Japan	0.85237	5	Ireland	58.8465
6	France	0.84814	6	Israel	58.7666
7	Finland	0.84110	7	Japan	58.6854
8	Israel	0.82782	8	Finland	58.5632
9	Sweden	0.82603	9	Sweden	58.4986
10	Austria	0.81834	10	France	58.4445
11	South Korea	0.81251	11	Netherlands	58.3931
12	Netherlands	0.79620	12	South Korea	58.1906
13	Ireland	0.79595	13	United States	58.1685
14	United States	0.78756	14	Greece	57.6726
15	United Kingdom	0.78369	15	United Kingdom	57.6301
16	Italy	0.78223	16	Denmark	57.5464
17	Greece	0.77738	17	Italy	57.3957
18	Spain	0.77302	18	Spain	57.3655
19	Denmark	0.77230	19	Austria	57.3463
20	Slovenia	0.77229	20	Czech Republic	56.9904
21	Portugal	0.77201	21	Slovenia	56.6348
22	Czech Republic	0.77170	22	Malta	56.1843
23	Cyprus	0.76857	23	Portugal	55.8007
24	Luxembourg	0.76248	24	Cyprus	55.7636
25	Chile	0.76149	25	Luxembourg	55.7535
26	Malta	0.75931	26	Chile	55.5452
27	Costa Rica	0.75243	27	Slovakia	55.2117
28	Brunei Darussalam	0.74769	28	Estonia	55.1126
29	Slovakia	0.73562	29	Poland	55.0607
30	Poland	0.73551	30	Lithuania	54.6504
31	Estonia	0.73548	31	Hungary	54.6273

32	Lithuania	0.73254	32	Brunei Darussalam	54.3089
33	Croatia	0.72985	33	Argentina	54.1389
34	Argentina	0.72963	34	Uruguay	53.9201
35	Uruguay	0.72898	35	Costa Rica	53.7812
36	Hungary	0.72886	36	Bahrain	53.7402
37	Panama	0.72831	37	Croatia	53.6746
38	Malaysia	0.71958	38	Latvia	53.5542
39	Bahrain	0.71867	39	Panama	53.1877
40	Latvia	0.71433	40	Malaysia	52.6534
41	Ecuador	0.70750	41	Bosnia and Herzegovina	52.5904
42	Mexico	0.70476	42	Bulgaria	52.5300
43	Bosnia and Herzegovina	0.69985	43	Mexico	52.3198
44	Bulgaria	0.69834	44	Ecuador	52.0422
45	Peru	0.69485	45	Peru	52.0041
46	Brazil	0.69244	46	Ukraine	51.7983
47	Armenia	0.69019	47	Armenia	51.6044
48	Colombia	0.68479	48	Qatar	51.4680
49	Tunisia	0.68385	49	Tunisia	51.1463
50	Venezuela	0.67498	50	Brazil	51.1128
51	Qatar	0.67452	51	Jordan	50.6354
52	Ukraine	0.67398	52	Colombia	50.4288
53	Philippines	0.67363	53	Philippines	50.1997
54	Jordan	0.67068	54	Saudi Arabia	50.0211
55	Paraguay	0.66369	55	Venezuela	50.0175
56	China	0.66283	56	Algeria	49.6628
57	Saudi Arabia	0.65925	57	Paraguay	49.6511
58	Algeria	0.64935	58	China	49.4022
59	Honduras	0.64414	59	Iran	49.2537
60	Iran	0.63915	60	Moldova	49.0966
61	El Salvador	0.63833	61	El Salvador	49.0601
62	Turkey	0.63746	62	Russia	48.8639
63	Trinidad and Tobago	0.63606	63	Trinidad and Tobago	48.7595
64	Moldova	0.62010	64	Kazakhstan	48.4834
65	Russia	0.62003	65	Turkey	48.4652
66	Indonesia	0.61709	66	Honduras	48.3524
67	Guatemala	0.61224	67	Bolivia	47.9357

68	Bolivia	0.60794	68	Indonesia	47.5274
69	Mongolia	0.60643	69	Mongolia	47.2246
70	Kazakhstan	0.59485	70	Tajikistan	46.7772
71	Morocco	0.57983	71	Guatemala	45.8716
72	Tajikistan	0.55936	72	Morocco	45.4151
73	Pakistan	0.53821	73	Togo	42.7150
74	Togo	0.52843	74	Pakistan	42.4818
75	India	0.48745	75	India	42.2151
76	Bangladesh	0.48236	76	Namibia	42.2034
77	Myanmar	0.46823	77	Bangladesh	41.7758
78	Cambodia	0.43702	78	Cambodia	41.3812
79	Namibia	0.41400	79	Myanmar	40.8534
80	Yemen	0.40035	80	Ghana	39.4355
81	Ghana	0.36921	81	Yemen	39.2093
82	Senegal	0.33812	82	Kenya	38.9111
83	Kenya	0.31422	83	Senegal	37.5300
84	Cameroon	0.28452	84	Cameroon	36.4265

85	Albania	N/A	85	Albania	N/A
86	Angola	N/A	86	Angola	N/A
87	Azerbaijan	N/A	87	Azerbaijan	N/A
88	Belarus	N/A	88	Belarus	N/A
89	Canada	N/A	89	Canada	N/A
90	Congo, Rep.	N/A	90	Congo, Rep.	N/A
91	Cote d'Ivoire	N/A	91	Cote d'Ivoire	N/A
92	Cuba	N/A	92	Cuba	N/A
93	Egypt	N/A	93	Egypt	N/A
94	Eritrea	N/A	94	Eritrea	N/A
95	Ethiopia	N/A	95	Ethiopia	N/A
96	Gabon	N/A	96	Gabon	N/A
97	Georgia	N/A	97	Georgia	N/A
98	Germany	N/A	98	Germany	N/A
99	Jamaica	N/A	99	Jamaica	N/A
100	Lebanon	N/A	100	Lebanon	N/A
101	Libya	N/A	101	Libya	N/A
102	Mozambique	N/A	102	Mozambique	N/A

103	Nepal	N/A	103	Nepal	N/A
104	Oman	N/A	104	Oman	N/A
105	Singapore	N/A	105	Singapore	N/A
106	Sri Lanka	N/A	106	Sri Lanka	N/A
107	Sudan	N/A	107	Sudan	N/A
108	Syria	N/A	108	Syria	N/A
109	Tanzania	N/A	109	Tanzania	N/A
110	Thailand	N/A	110	Thailand	N/A
111	United Arab Emirates	N/A	111	United Arab Emirates	N/A
112	Vietnam	N/A	112	Vietnam	N/A
113	Zambia	N/A	113	Zambia	N/A
114	Zimbabwe	N/A	114	Zimbabwe	N/A