The Dimensionality of the Measures in the Human Development Index

David Canning‡ , Declan French†, and Michael Moore†

† *UKCRC Centre of Excellence for Public Health, School of Management, Queens University, Belfast BT7 1NN. (email - declan.french@qub.ac.uk)*

‡ *Department of Global Health and Population, Harvard School of Public Health, Harvard University,*

February 2011

Abstract

The Human Development Index compresses four measures of development into a single indicator. Using non-parametric methods we show that the four measures have dimension three and the index therefore loses information. The education measures, literacy and school enrolment, have dimension less than two, but the three variables school enrolment, income per capita, and life expectancy, appear to fill a three dimensional space.

JEL Classification*:* I32, O10, O15

Keywords*: development, well-being, dimension, measure, indicator.*

1 Introduction

The Human Development Report 1990 (United Nations Development Program 1991) identified three important components of human development: longevity, knowledge, and the standard of living. Taking four indicators, life expectancy, literacy and school enrolment rates, and income per capita as measures of these components it puts forward the Human Development Index (HDI), based on relative performance of countries on these measures, as a summary measure of development.

The Human Development Index has been criticized on two main grounds; it leaves out important measures of development, and the aggregation into a single composite index is misleading (Kovacevic (2010)). In this paper we take the components in the human Development Index as given and focus on the issue of aggregation (French, Moore and Canning (2010) investigate the aggregation issue if we admit larger number of development indicators). The components of development can be aggregated into a single index of wellbeing if we have a well defined social welfare function. However, it is difficult to construct such a welfare function since it involves interpersonal comparisons of utility. The Human Development Index does not try to construct a welfare measure but is a simple average of the three components. A justification for this approach is that even though the three components measure different things, they move together, and a single index is a good proxy for the overall level of human development in a country.

While countries could have arbitrary values of each of development component, in practice, a country's position could be completely determined by a single parameter. For example, we might think of health and education as being determined by income level, or income levels could be determined by human capital in the form of health and education. For example Preston (1975) observes a nonlinear relationship between income levels and life expectancy across countries a relationship that seems to continue to hold though the direction of causality may be unclear (Bloom and Canning, 2007). Ranis, Stewart, and Ramirez (2000) find linkages between many measures of development.

A number of studies have used correlation structures and principle component analysis to study this issue for the components of the Human Development Index (for example, Cahill (2005), Ivanova, Arcelus and Srinivasan (1999), and Saisana, Tarantoorbakhshola and Saltelli, (2005)). These linear techniques suggest that the data has a single common component but there is no consensus in these papers on the question of whether the degree of correlation in the variables is sufficiently high to allow compression of the data into a single index number of if significant information is lost in doing this.

In this paper we adopt an alternative approach. We seek to determine the dimension of this subspace spanned by the four variables that make up the Human Development Index. Conceptually the dimension of a space is the minimum number of coordinates needed to identify a point in the space. We employ a non-parametric estimator of the dimensionality of the data using techniques developed by Grassberger et al (1983) and Takens (1985). Results based on this non-parametric approach indicate the four measures used in construction of the human development index have dimension three. The literacy and school enrolment measures have dimension less than one showing that they are closely linked, but school enrolment, taken together with life expectancy and income per capita, seems to have dimension three.

The non parametric approach allows for non-linear data structures and can capture structural relationships that are hidden from models that assume a linear relationship. We do find a non-linear structure in the relationship between school enrolment and literacy (literacy rates rise with school enrolment at first but then are bounded at 100% for high levels of enrolment). While the other measures are correlated the large variation around the relationship mean that the data fills a three dimensional space. Our results support the idea that development is fundamentally multidimensional and we need multidimensional measures to describe development, using a “dashboard” of indicators rather than a single index (The *Commission on the Measurement of Economic Performance and Social Progress* (2009) , Alkire (2002)).

Section 2 describes our analysis and in the conclusion we discuss the reasons for the different estimates of the dimensionality of the data.

2. Analysis

We examine the four variables that are used in the construction of the Human Development Index: life expectancy at birth, the adult literacy rate, combined (primary, secondary and tertiary) school enrolment rates, and real GDP per capita. Data is from HDI Trends and Indicators (1980-2007) in United Nations Development Program Report (2009) and covers 182 countries or the year 2007. All countries have complete data (in some cases the report imputes data when data is missing in order to have complete data and an index for every country). Table 1 shows descriptive statistics while Table 2 shows a correlation matrix for the four data series.

In constructing the Human Development Index the data are transformed, first by using the log of real GDP per capita and secondly by rescaling each variable by measuring its position as a proportion of the range between that variable’s minimum and maximum values. An average of these rescaled variables is used as the index. In our analysis we use the raw data; a strength of our non-parametric approach is that it is robust to functional transformations of the data.

We endeavor to estimate the dimensionality of the data directly using a nonparametric method. If data points are uniformly distributed on a space of dimension d the number of points in a ball of radius R will be proportional to R^d . If follows that as we increase the radius around a particular data point we will relatively find more neighbors when the radius is larger if the space spans a larger dimension, and that the average distance between a point and its nearest neighbors increases with the dimension of the space (Bhattacharyya and Chakrabarti (2008) provide an exact formula a simple proof). The Takens estimator (1985) uses this intuition. It examines the distribution of distances R_i between pairs of points for all pairs with distances less than some threshold R_0 and is based on the fact that $S_i = -\log(R_i / R_0)$ for $R_i < R_0$ has a probability density given by the exponential function with parameter d where d is the dimension of the space. From the properties of the exponential distribution, it then follows that the maximum likelihood estimator of d is given by the reciprocal of the mean of S_i, $\hat{d} = \overline{S}_i^{-1}$ and the standard error can be estimated by $\hat{\sigma} = \hat{d} / \sqrt{N^*}$ where N* is the number of distances between pairs of points with lengths $R_i < R_0$. An estimate, $\hat{d}$, and a confidence interval can be then determined for the dataset for a given choice of boundary radius R_0.

If R_0 is small, the estimate depends on only a few distances, and has a very large standard error. However as R_0 gets very large the estimate is affected by the presence of empty space on the 'edge' of the dataset. To address this we follow the procedure suggested by Smith (1997) which involves using for the largest R_0 for which the estimate lies inside the confidence interval of estimates for any smaller R_0. Using a small value of R_0 not only gives a large standard error but tends to give a large estimate of dimension since any random noise will make the data fill the space available in the local neighbourhood of any relationship. However when R_0 is larger than the range of the support of the noise the effect of the noise on the dimension estimate will be small.

In Figure 1 we plot the estimate of dimension d, ands the 99% confidence interval against the number of data pairs N^* used to calculate the estimate. The number of data pairs

$N*$ increases monotonically in the threshold distance R_0 chosen. The Smith estimate is based on the largest R_0 which gives an estimate that is within the 99% confidence interval for any smaller R_0. This approach gives us an estimate of the dimension 2.93 with a 95% confidence interval of 2.72 to 3.14. Even for much larger values of $N*$ and R_0 the dimension estimate remains at around 3.

In Table 3 we give results for all four variables that underlie the human development index and compare them with estimates of dimension for a number of pairs of variables. With two variables we can have dimension up to two. In row 2 of Table 3 we have estimate the dimension spanned by the data on enrolment and literacy. The dimension estimate is very low, 0.38 and even the upper 95 percent confidence limit for the dimension is less than one. Both enrolment rates and literacy rates in the HDI data set are bounded (literacy rates at 99% by convention and enrolment rates at 100%). A number of countries have attained these boundary levels of enrolment and literacy. These countries are therefore at exactly the same point in the two dimensional space; if all points were at this boundary we would have a set of dimension zero. In addition, we have a number of countries that have attained 99% literacy with less than 100% enrolment. These countries have data pints that lie on a one dimensional line with literacy at 99 %. These clusters of data points at one location and along a one dimensional line leads to a very low estimated dimension for the two variables.

On the other hand, data on school enrolment and income (measured as GDP per capita) fill the two dimensional space. Our estimate of the dimension of the data in this case, given in Row 3 of Table 3 is 1.82, though the 95% confidence interval includes dimension 2. Similarly when we measure the dimension spanned by school enrolment and life expectancy in Row 4, and of life expectancy and GDP per capita in Row 5 we find point estimates that are close to 2 and have 95% confidence intervals that include dimension 2.

Our results differ from the maximum likelihood estimate based on linear principal component analysis which indicates one common component. However this component explains only a fraction of the variation in the data. The non-parametric estimate of dimension shows that the remaining unexplained variation is large enough to fill a three dimensional space.

3. Conclusion

This paper introduces a non-parametric technique to estimate the dimension of development. Our results indicate that the four components of the United Nations Human development Index have dimension three. The literacy and school enrolment data is of low dimension due to countries being clustered at the boundary where enrolment and literacy are at, or close to, 100%. However other pairs of indicators seem to have dimension two and the overall data set containing an income, a health, and an education measure has dimension three. This implies that compressing the data into a single index number loses information.

Bibliography

Alkire S. (2002), 'Dimensions of Human Development', *World Development*, Vol. 30, No.2, pp 181-205.

Bhattacharyya, P. and Chakrabarti B.K. (2008), 'The mean distance to the *n*th neighbour in a uniform distribution of random points: an application of probability theory', European Journal of Physics, Vol. 29**,** No. 3, pp. 639-645.

Bloom, D. E., and Canning, D. (2007), 'The Preston Curve 30 Years On: Still Sparking Fires', *International Journal of Epidemiology*, Vol. 36, No. 3, pp. 498-499.

Cahill, M. (2005), 'Is the Human Development Index Redundant?', *Eastern Economic Journal*, Vol. 31, No. 1, pp. 1-5.

CMEPSP (2009), *Report of the Commission on the Measurement of Economic Performance and Social Progress*, www.stiglitz-sen-fitoussi.fr/en/index.htm.

French, D., Moore, M. and Canning, D. (2010), 'Is Human Development Multidimensional', Finance and Economics Research Group Working Paper, Queen's University of Belfast.

Grassberger, P. and Procaccia, I., (1983), 'Measuring the strangeness of strange attractors', *Physica D*, Vol. 9, pp. 189-208.

Ivanova, I., Arcelus, F.J. and G. Srinivasan. (1999), 'An assessment of the measurement properties of the human development index', *Social indicators research*, Vol. 46, pp 157–179.

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

Preston, S. (1975), 'The Changing Relation between Mortality and Level of Economic Development', *Population Studies*, Vol. 29, No. 2, pp. 231-248.

Ranis G., Stewart F., and Ramirez A. (2000), 'Economic Growth and Human Development', *World Development,* Vol. 28, No 2, pp 197-219.

Saisana, M., Tarantoorbakhshola, S., and Saltelli, A. (2005). 'Uncertainty and Sensitivity Techniques as Tools for the Analysis and Validation of Composite Indicators', *Journal of the Royal Statistical Society A*, Vol. 168, No. 2, pp 1-17.

Smith, L. A. (1997), "The maintenance of uncertainty", in G. Cini Castagnoli and A. Provenzale, eds., *Past and Present Variability in the Solar-Terrestial System: Measurement, Data Analysis and Theoretical Models*, volume CXXXIII of International School of Physics\ Enrico Fermi, pp. 177-246, Bologna, 1997. Il Nuovo Cimento.

Takens, F. (1985), 'On the numerical determination of the dimension of an attractor,' in *Dynamical Systems and Bifurcations*, Groningen, 1985, B. L. J. Braaksma, H. W. Broer, and F. Takens, eds., Vol. 1125 of Lecture Notes in Mathematics, Springer-Verlag, Berlin.

United Nations Development Program (1991), *Human Development Report 1990*, Oxford University Press, New York.

United Nations Development Programme (2008), *Human Development Indices – A Statistical Update*, United Nations, New York. hdr.undp.org/en/statistics/data/hdi2008/.

Table 1

Descriptive Statistics for the Components of the human Development Index, 2007

	Mean	Standard Deviation	Minimum	Maximum
Literacy Rate %	83.7	19.0	26.2	99
School Enrolment Rate %	72.1	17.0	25.5	100
Real GDP per Capita, $	13482	15827	298	85382
Life Expectancy years	68.5	10.2	43.6	82.7

For 182 countries, from United Nations Development Programme (2008).

Table 2

Correlation Matrix

	Literacy Rate	School Enrolment Rate	Real GDP per Capita	Life Expectancy
Literacy Rate	1			
School Enrolment Rate	0.809	1		
Real GDP per Capita	0.490	0.587	1	
Life Expectancy	0.740	0.773	0.603	1

Table 3

Non Parametric Dimension Estimate

	Number of Variables	Estimated Dimension	95 % lower Confidence Interval	95 % upper Confidence Interval per
All four Measures	4	2.93	2.72	3.14
Literacy and School Enrolment	2	0.38	0.27	0.50
School Enrolment and GDP per capita	2	1.82	1.59	2.06
School Enrolment and Life Expectancy	2	1.88	1.52	2.24
GDP per Capita and Life Expectancy	2	2.05	1.70	2.40

Estimates constructed using Takens' Maximum Likelihood Estimator with Smith's Method of choosing the threshold.

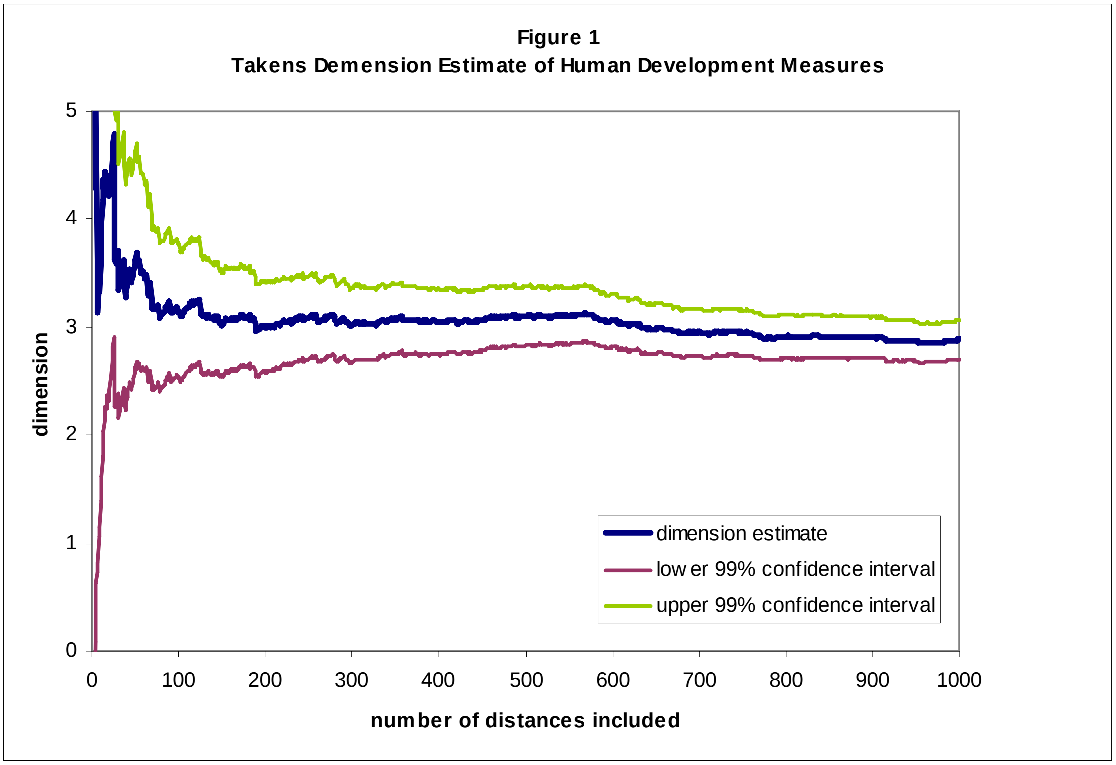
Figure 1
Takens Demension Estimate of Human Development Measures
dimension
0
1
2
3
4
5
0
100
200
300
400
500
600
700
800
900
1000
number of distances included
dimension estimate
low er 99% confidence interval
upper 99% confidence interval