Session Number: 8B
Session Title: Issues in Income Distribution
Paper Number: 2.5
Session Organizer: Edward Wolff, NYU, USA

Paper Prepared for the 26th General Conference of The International Association for Research in Income and Wealth, Cracow, Poland, 27t August to 2 September 2000

TITLE: THE CHOICE OF PRINCIPAL VARIABLES FOR COMPUTING SOME MEASURES OF HUMAN WELL-BEING

AUTHORS: TOMSON OGWANG AND ABDELLA ABDOU

For additional information please contact:
Tomson Ogwang
Economics Program
University of Northern British Columbia
3333 University Way
Prince George, British Columbia
V2N 4Z9
Email: ogwang@unbc.ca
Fax: (250) 960 5544
Telephone: (250) 960 6485

This paper is placed on the following websites: ***www.stat.gov.pl***
www.econ.nyu.edu/dept/iariw

Abstract

The present paper investigates the choice of principal variables for computing three measures of human well-being, namely, the Human Development Index (HDI), the Gender-Related Development Index (GDI), and the Gender Empowerment Measure (GEM). To this end, the principal components variable selection strategy considered by Jolliffee (1972, 1973) is applied to data from the 1999 Human Development Report. The empirical results show that there is statistical justification for selecting only one of the three components of each indicator. In the case of the HDI, the life expectancy index turns out to be the best choice; in the case of the GDI the best choice is the equally distributed educational attainment index; and in the case of the GEM the best choice is the index of parliamentary representation. Accordingly, it seems reasonable to compute simpler HDI, GDI and GEM based solely on the selected indicators without loss of too much information. Also, while our results support the current practice of equally weighting the three components of HDI and GDI, this is not the case for the GEM.

1. Introduction

As evidenced by a spate of recent papers on the subject, there is considerable interest in the concept and measurement of human development. See, for example, Hopkins (1991), McGillivray (1991), Dasgupta and Weale (1992), McGillivray and White (1993), Srivastavan (1994), Streeten (1994), and Pillarisetti and McGillivray (1998), among others. There are several plausible explanations for this. First, it is now widely recognized that GDP (or GNP) per capita is not a fully satisfactory measure of human well-being and that composite indicators, that incorporate more than one attribute of human well-being, are conceptually more appealing. Second, the popularization of the Basic Human Needs (BHN) approach to development in the 1970s led to the search for BHN performance indicators, which have facilitated the construction of human development indicators.[1] The BHN approach emphasizes the satisfaction of basic needs such as good nutrition, health, education, clothing and sanitation. In contrast, the traditional growth strategies emphasize increasing GDP (or GNP) per capita. Third, the annual publication of the Human Development Report (HDR) by the United Nations Development Programme (UNDP) since 1990 has raised awareness of and focused debate on the importance of the concept of human development. In fact, over the past decade the UNDP has taken the lead in conceptualizing and measuring the conditions of human development.

In the inaugural issue of the HDR, the UNDP(1990) proposed the Human Development Index (HDI) which is a composite index comprising of life expectancy, adult literacy, and real GDP per capita adjusted for purchasing power parity. The index ranges from zero to one with

[1] Some measures of BHN performance proposed by Hicks and Str incorporated in the HDI.

higher values signifying higher levels of human development. The HDI is rationalized on the grounds that human beings strive to lead long healthy lives (captured by life expectancy), to have a descent standard living (captured by real GDP per capita), and to acquire knowledge (captured by adult literacy), in which case the focus of development policy should be to expand the choices and capabilities of people in these areas. The HDI is increasingly being used as a basis for setting human development goals of many countries and more than 100 countries are already constructing national or sub-national HDIs.

In response to some of the criticisms of the earlier versions of HDI, the UNDP has continued to refine the underlying analytical framework. To this end, several improvements have been made in the methodology and quality of data used for constructing the HDI. The UNDP (1995) introduced two additional measures of human development, namely, the Gender-Related Development index (GDI) and the Gender Empowerment Measure (GEM). As we shall see in section 2 below, the GDI uses the same dimensions as the HDI except for the fact that the former takes into account gender disparities in human development. Specifically, the GDI is constructed such that countries that exhibit greater gender disparities in human development are penalized by achieving lower GDI values relative to their HDI counterparts. The GEM is a measure of the degree of womens' participation in the political, economic, and professional activities. It must be pointed out that the GDI and GEM, like the HDI, lie between zero and one with greater values signifying higher levels of human development.

The HDI, GDI and GEM have the common characteristic of incorporating more than one attribute of human well-being. Although the addition of extra attributes may be appealing on conceptual grounds, it also introduces new complications with respect to the assignment of weights to the attributes, and the added time and monetary costs of obtaining information on all attributes, among others.[2] Even the proponents of the HDI, GDI and GEM concede that it is impossible to come up with a comprehensive set of human development indicators. See, for example, UNDP (1999, p. 127). In fact, as pointed out by UNDP (1994, p. 91), more indicators is not necessarily better in the sense that there might be an overlap among some indicators as is the case with infant mortality which is already reflected in life expectancy. The fact that it is impossible to come up with a single composite indicator of human development that incorporates all attributes of human well-being raises the question as to whether or not there is statistical justification for including all dimensions of a particular composite indicator of human well-being. In our opinion, it is important to incorporate only attributes that add new information. Accordingly, the most parsimonious, yet informative, composite indicators of human development are recommended.

In a recent study of the HDI, Ogwang (1994) found that there is statistical justification for retaining only one of the three components of HDI. It must be noted that Ogwang's study was based on data for an earlier version of HDI obtained from the 1991 HDR. Given recent refinements in the methodology for computing the HDI as well as improvements in the quality of the underlying data, it seems reasonable to re-examine this issue using the most recently available and more reliable data for the HDI. Furthermore, to the best of our knowledge, no empirical studies have hitherto examined the issue of choice of principal variables for computing the GDI and GEM. The present paper fills in the gap by examining the choice of principal variables for

[2] For a discussion of the difficulties in measuring the literacy component of HDI, see Hopkins (1992, p. 1471).

computing HDI, GDI, and GEM. To this end, the principal components variable selection strategy considered by Jolliffe (1972,1973) is used to analyze data from the 1999 HDR.

The format of the rest of the paper is as follows. In section 2, the methodologies for constructing the HDI, GDI and GEM are briefly discussed. In section 3, the variable selection strategy employed is briefly described. In the penultimate section, the empirical results are presented. The concluding remarks are made in the final section.

2. Construction of the HDI, GDI and GEM

The most recent versions of HDI, GDI and GEM, reported in the 1999 HDR, are each computed as simple arithmetic means of three sub-components. The HDI is computed as a simple arithmetic mean of the life expectancy index (LEI), educational attainment index (EAI), and an index of real GDP per capita adjusted for purchasing power parity (GDPI). Prior to computing the mean, each of the three components is scaled to lie between zero and one. For further details concerning the construction of the most recent version of HDI, see UNDP (1999, pp. 159-160).

The GDI is computed as a simple arithmetic mean of an equally distributed life expectancy index (ELEI), an equally distributed educational attainment index (EEAI), and an equally distributed income index (EGDPI). Prior to computing the mean, each index is scaled to lie between zero and one. The equally distributed components take into account differencial life expectancies, educational attainments and incomes of men and women as well as their differential population shares. For further details concerning construction of the most recent version of GDI, see UNDP (1999, pp. 160-161)

The GEM is computed as a simple arithmetic mean of an index of parliamentary representation (PRI), index of administrative and managerial, and professional and technical positions (AMPTPI), and an equally distributed income index (EGDPI). Prior to computing the mean, each index is scaled to lie between zero and one. For further details concerning construction of the most recent version of GEM, see UNDP (1999, pp. 161-162).

Note that taking the arithmetic mean of the three components of each indicator amounts to assigning equal weights to these components, implying that they are equally important in that indicator. In the sections below, we will examine the issue of which component best represents the three components of each of the three human development indicators (HDI, GDI and GEM).

3. The selection strategy

Principal components analysis (PCA) technique has traditiona
transform a large set of correlated variables into a smaller set of
called the principal components, that account for most of the va
of variables. Since the principal components are linear combinatio
with mathematically determined characteristic vectors of the co
matrix of original variables as weights, it can be argued that PCA r
arbitrary choice of weighting scheme. The first principal compon
proportion of variation in the original set of variables. The secon
captures the largest proportion of variation which is not accou
component, and so on. For a set of k variables, the maximum numbe
which can be extracted is equal to k. However, if the k variables ar

few principal components capture most of the variation. The prop
attributed to a particular principal component is obtained by div
characteristic root by the sum of all the characteristic roots. Du
good introduction to PCA.

In principle, the high inter-correlations among the components of composite indicators could be exploited to construct new composite indicators which are, in fact, first principal components of the original components and then scaling them to lie within the desired range. A drawback of this approach is that information on all the components is required since each principal component is a linear combination of all the constituent variables.[3]

Recently, PCA has also served as a useful tool for selecting a few variables among a wider set of correlated variables. The technique used in the present paper is PCA met
Jolliffe (1972, p. 164; 1973, p. 22), which is recommended by Jolliffee (1972
retention of best subsets of variables is considered important, wh
the present study. Essentially, the method involves first selecting
highest correlation with the first principal component. This is f
which has the highest correlation with the second principal co
stage selection process is limited only to those variables that are
selection process. This procedure is continued until the required
selected. Jolliffe (1972, p. 171) recommended that the number of vari
equal to the number of principal components associated with cha

[3] See, for example, Ram (1982) and UNDP (1993, pp. 109-110).

correlation matrices of variables which are greater than 0.70. If t
variables is used, the number of variables to be selected is equal to
characteristic roots of the covariance matrix which are greater
average of the characteristic roots associated with the covarian
principal components literature. See Dunteman (1989, p. 22) for furt

4. Empirical Results

The empirical results presented in this section are based on t
available data taken from the 1999 HDR. The data for the HDI relate
those for the GDI and GEM pertain to 143 and 102 countries, respec
number of countries is dictated by the availability of data as repo

To provide some insights into the degree of association amo
of the three components of each of HDI, GDI, and GEM, the covaria
moment correlation coefficients and the Spearman's rank correl
reported in Table 1. Evidence of strong positive correlation amon
each human development indicator is apparent from the table. Th
are exhibited by the three GEM components.

Table 1 here

Since the selection strategy employed here requires knowled
and characteristic vectors of the covariance matrices of the thr
indicator, the covariances and the associated characteristic roo

HDI components are reported in Table 2. Similar data for the GDI ar and Table 4, respectively.[4] It can be seen from the entries in Table 2 component accounts for approximately 85 percent of the variatio HDI. The first two principal components account for approximate in the three variables. Clearly, the first principal component cont information embedded in the three components of HDI.

The components of the characteristic vector associated wit component of the three HDI components show that the first prin could be computed as 0.564 (LEI) + 0.620 (EAI) + 0.546 (GDPI). However, thi computing all the three components, which can be expensive to co the components of the characteristic vector associated with the do not sum to unity, the first principal component HDI series will one without further transformation.[5] Another notable feature o principal component weights attached to the three HDI compone which is consistent with the equal weighting scheme adopted by

---

[4]The present study uses the covariance matrix of variables rat matrix since the units of measurement of the three components o reasonably commensurate, as suggested in the principal componen Morrison, 1967, p. 223).

[5] For the weights to sum to unity, the weight attached to eac divided by the sum of all the weights in each case.

of the HDI in the 1999 HDR. It is also easy to verify from the entries
characteristic root is greater than 0.708, suggesting that only o
retained for purposes of computing a simpler HDI.

Table 2 here

Since the decision as to which of the three variables should
information on the correlations (or loadings) between the varia
components, these correlations are also reported in Table 2. It can
the table that the life expectancy index (LEI) has the highest corr
principal component. Accordingly, it can be concluded, based on
selection criterion, that the life expectancy index (LEI) is the best
series based solely on the LEI would be cheaper to compute withou
information. In fact, the product moment correlation and rank
between the HDI and LEI are 0.934 and 0.944, respectively, which are
near perfect positive correlation.

Table 3 shows the results of PCA of the three GDI components.
table are in order. First, the first principal component accounts f
of the variation in the three components of GDI. The first two pri
for approximately 94 percent of the variation in the three variabl
principal component contains most of the statistical informatio
components. Second, the components of the characteristic vecto
principal component of the three GDI components show that the

GDI series could be computed as 0.542 (ELEI) + 0.644 (EEAI) + 0.539 (EGDPI). necessitates data for computing all the three components, which Furthermore, since the components of the characteristic vector principal component do not sum to unity, the first principal com between zero and one without further transformation. Third, th weights attached to the three GDI components are approximately with the equal weighting scheme adopted by the UNDP in the esti HDR. Fourth, it is also easy to verify from the entries in Table 3 tha root is greater than 0.708 suggesting that only one of the three purposes of computing a simpler GDI. An inspection of the correla and the principal components and application of the aforementio indicates that the selected variable should be the equally distrib index (EEAI). Thus, a simpler GDI series based solely on EEAI would be loss of too much information. In fact, the product moment corre coefficients between the GDI and EEAI are 0.924 and 0.907, respective one indicating near perfect positive correlation. We also note th the sole indicator for constructing the GDI in the 1999 HDR, the would have increased from 143 to 145.

Table 3 here

Table 4 shows the results of PCA of the three GEM components table are worthy of mention. First, the first principal component

65 percent of the variation in the three GEM components. The first account for approximately 85 percent of the variation in the thre with the HDI and GDI, the first principal component contains mos information embedded in the three GEM components. Second, the characteristic vector associated with the first principal compon components show that the first principal component GEM series c + 0.388 (AMPTPI) + 0.481 (EGDPI). However, this also necessitates data fo components, which can be expensive to collect. Furthermore, sinc characteristic vector associated with the first principal compon first principal component GEM series will not lie between zero and transformation. Third, unlike in the cases of the HDI and GDI, the weights attached to the three GEM components are not approxim conformity with the equal weighting scheme adopted by the UND GEM in the 1999 HDR. Fourth, it is also easy to verify from the entri characteristic root is greater than 0.708 suggesting that only on retained for purposes of computing a simpler GEM. An inspection o the variables and the principal components and application of th criterion shows that the selected variable should be the index o representation (PRI). Thus, a simpler GEM series based solely on the P compute without loss of too much information. In fact, the prod rank correlation coefficients between the GEM and PRI are 0.857 an

are very close to one indicating near perfect positive correlation
PRI been used as the sole indicator for constructing the GEM in t
countries ranked would have increased from 102 to 164.

Table 4 here

5. Concluding Remarks

In this paper, the principal components variable selection str
Jolliffe (1972, 1973) is used to investigate the choice of principal var
human development indicators, namely, the HDI, GDI and GEM usin
HDR. The life expectancy index (LEI) turns out to be the best choice
simplified HDI; in the case of the GDI, the equally distributed educ
(EEAI) turns out to be the best choice; and in the case of the GEM t
representation (PRI) is the best choice. Furthermore, the correlati
indicator and the best variable for constructing that indicator
Accordingly, it seems reasonable to compute simpler HDI, GDI and
selected indicators without loss of too much information. Our resu
the current practice of equally weighting the three GEM components. Finally, it must be pointed out that our discussion concerning the choice of indicators is based purely on the analysis of the internal structures of the data and does not preclude the use of other criteria for inclusion of other

variables to these indicators.

References

Dasgupta, R. and M. Weale (1992), "On Measuring the Quality of Life," W

Dunteman, G.H. (1989), Principal Components Analysis, Sage Publicati

Hicks, N. and P. Streteen (1979), "Indicators of Development: The Sear Yardstick," World Development, 7, 567-580.

Hopkins, M. (1991), "Human Development Revisited: A New UNDP Report 1473.

Jolliffe, I.T. (1972), "Discarding Variables in a Principal Component An Applied Statistics, 21, 160-173.

Jolliffe, I.T. (1973), "Discarding Variables in a Principal Component An Statistics, 22, 21-31.

McGillivray, M. (1991), "The Human Development Index: Yet Another Re Development Indicator?," World Development, 19, 1461-1468.

McGillivray, M. and H. White (1993), "Measuring Development? The UN Index," Journal of International Development, 5, 193-205.

Morrison, D.F. (1967), Multivariate Statistical Methods, McGraw-Hill

Ogwang, T. (1994), The Choice of Principal Variables for Computing t Index," World Development, 22, 2011-2014.

Pillarisetti, J.R. and M. McGillivray (1998), "Human Development and G Methodological and Measurement Issues," Development Policy Revi

Ram, R. (1982), "Composite Indices of Physical Quality of Life, Basic Nee Principal Component Representation," Journal of Development Eco

Rao, V.V.B. (1991), "Human Development Report 1990: Review and Assessm 1451-1460.

Srivastavan, T. N. (1994), Human development,: A New Paradigm or Rein

American Economic Review Papers and Proceedings, 84, 238-243.

Streteen, P. (1994), Human Development: Means and Ends, "American Ec
Proceedings, 84, 238-243.

United Nations Development Programme (1990), Human Development Report 1990, New York, Oxford University Press.

United Nations Development Programme (1993), Human Development Report 1993, New York, Oxford University Press.

United Nations Development Programme (1995), Human Development Report 1995, New York, Oxford University Press.

United Nations Development Programme (1999), Human Development Report 1999, New York, Oxford University Press.

Table 1
Covariances and correlations among the three components of HDI, GDI and GEM

Development indicator		Covariance	Pearson's correlation	Spearman's correlation
HDI	LEI and EAI	0.029	0.779	0.772
	LEI and GDPI	0.027	0.809	0.858
	EAI and GDPI	0.027	0.73	0.733
GDI	ELEI and EEAI	0.03	0.776	0.788
	ELEI and EGDPI	0.027	0.811	0.872
	EEAI and EGDPI	0.03	0.761	0.787
GEM	PRI and AMPTPI	0.018	0.477	0.493
	PRI and EGDPI	0.019	0.438	0.365
	EGDPI and AMPTPI	0.01	0.319	0.415

Table 2

Results of PCA of the three HDI components using their covariance matrix

	Principal component		
	First	Second	Third
Characteristic root	0.090849	0.01024	0.00605
Percentage of variance explained individually	84.8	9.6	5.6
Percentage of variance explained cumulatively	84.8	94.4	100
Characteristic vectors* LEI	0.564	-0.262	0.783
EAI	0.620	0.761	-0.192
GDPI	0.546	-0.594	-0.591
Correlation coefficient with* LEI	0.931	-0.145	0.334
EAI	0.922	0.380	-0.074
GDPI	0.908	-0.332	-0.254

* LEI, EAI, and GDPI denote the life expectancy index, the educational attainment index, and the index of real GDP per capita adjusted for purchasing power parity, respectively.

Table 3
Results of PCA of the three GDI components using their covariance matrix

	Principal component		
	First	Second	Third
Characteristic root	0.09656	0.01007	0.00631
Percentage of variance explained individually	85.5	8.9	5.6
Percentage of variance explained cumulatively	85.5	94.4	100
Characteristic vectors* ELEI	0.542	-0.388	-0.745
EEAI	0.644	0.761	0.072
EGDPI	0.539	-0.519	0.663
Correlation coefficient with* ELEI	0.922	-0.213	-0.324
EEAI	0.934	0.356	0.027
EGDPI	0.915	-0.284	0.287

* ELEI, EEAI, and EGDPI denote the equally distributed life expectancy index, the equally distributed educational attainment index, and the equally distributed income index, respectively.

Table 4

Results of PCA of the three GEM components using their covariance matrix

		Principal component		
		First	Second	Third
Characteristic root		0.075832	0.0241	0.017712
Percentage of variance explained individually		64.5	20.4	15.1
Percentage of variance explained cumulatively		64.5	84.9	100
Characteristic vectors*	PRI	0.786	-0.450	-0.424
	AMPTPI	0.318	-0.174	0.905
	EGDPI	0.481	0.876	-0.038
Correlation coefficient with*	PRI	0.924	-0.217	-0.241
	AMPTPI	0.615	-0.165	0.737
	EGDPI	0.698	0.716	-0.027

* PRI, AMPTPI, and EGDPI denote the index of parliamentary representation, the index of administrative and managerial, and technical positions, and the equally distributed income index, respectively.