Methodological Choices Encountered in the Construction of Composite Indices of Economic and Social Well-Being

Julia Salzman
Center for the Study of Living Standards
3/13/2003

Methodological Choices Encountered in the Construction of Composite Indices of Economic and Social Well-Being

Detailed Table of Contents:

List of Charts, Exhibits, Tables:

ABSTRACT:

In recent year a large number of composite indexes of economic and social well-being have been developed (Hagerty et. al., 2001). Unfortunately, the methodological issues associated with index construction have often been neglected or inadequately treated by index developers. The objective of this paper is to provide a comprehensive review of the methodological choices involved in the construction of indexes of economic and social well-being and the implications of the choices for the properties of the index. Building on a recent paper by Booysen (2002) and work done by the United National Development Program (e.g. Anand and Sen, 1994), the paper addresses issues related to the choice of functional form of variables, scaling issues, the aggregation operations, weighting schemes, and the choice between single and complementary composite indexes. A detailed typology of the issues addressed in the paper accompanies this abstract. The paper concludes with a list of recommendations for best-practice methodologies for composite index construction.

EXECUTIVE SUMMARY

The goal of this paper is to provide an overview of methodological choices encountered in the construction of composite indices of well-being. We develop a sequential typology of these choices that is demonstrated schematically in Exhibit 1: choosing a single or complementary approach to the index, selecting variables, determining the functional form of each variable, choosing a method of standardization, choosing an aggregation operation, and finally, determining a weighting scheme. By analyzing the interpretation of the main decisions made by popular indices of well-being, we hope to provide a basis on which more informed methodological decisions can be made. This analysis leads us some best practice conclusions.

The first choice encountered in index construction is the general form of the index: will it be a single composite, or a complementary composite. A single composite is a single aggregation of variables that are used in an index, whereas a complementary composite is comprised of two separate indices: a conglomerative index and a deprivational index. The complementary approach is due to Anand and Sen. A conglomerative index measures the overall well-being of a society, where increases in well-being of the best off can offset decreases in well-being of the worst off. In contrast, a deprivational index measures only the welfare of the worst off (Anand and Sen: 1997). Some of the variables used in a conglomerative index may also be used in its deprivational complement. The conglomerative approach has been adopted by the UNDP and applied in the complementary indices of the Human Development Index (HDI), a conglomerative index, and the Human Poverty Index (HPI), a deprivational index.

The next choice encountered is which variables to include in the index. This choice can be made by simply choosing data that an index constructor wants to include, or by first determining concepts that the developers seek to measure, such as inequality, and then finding data that represent this concept. In the latter case, it makes sense to think about the degree of correlation between different variables that measure the concept. If two variables are highly correlated, then inclusion of both of them is redundant. The correlation between two variables measures how redundant they are, and can be calculated easily. The level of a correlation can be used as a guide for which variables to include to reduce redundancy in data if this is a goal.

After variables have been picked, functional forms must be chosen. The functional form is a functional transformation that is applied to the raw data in order to represent the significance of marginal changes in its level. Most indices use the trivial linear functional form ($f(x)=x$) on all of the variables, de facto. However, this choice does not represent methodological consistency when the significance of marginal changes in the absolute value of a variable differ within the range of the variable taken on in the index. An example of the importance of non-linear functional forms is the per capita GDP in the Human Development Index (HDI). In this index, the log of the per capita GDP is taken because a $1000 US in income is much more significant to increases in quality of life at a low level of income than at a high one (see Chart 1). Index developers should examine whether marginal changes are significantly different within ranges taken on in a variable to determine whether a non-linear functional form should be applied. If marginal gains are more significant at a low level, the log is a good choice of functional form; if marginal gains are more significant at a high level, the power function (raising to the 2nd or 3rd power, for example) is a good choice.

Once functional forms associated to variables have been established, a uniform method of standardization should be considered. One choice is to use raw data and not standardize. This choice leads to many problems when an attempt is made to aggregate variables. Standardization methods allow standardized values to be compared meaningfully. Such methods are grouped into two categories: a focus on absolute value comparisons—both levels and changes—and a focus on percent changes over time. Three techniques to standardize absolute values of variables are reviewed: Linear Scaling Technique which linearly scales variables to a uniform range, ordinal response, where experts assign a score to each variable, and Gaussian normalization, or Z-score, in which the standardized variable is the number of standard deviations away from its mean. To focus on percent changes, the technique of normalization to base year is used, where each variable is divided by its level in the base year.

The choice of aggregation operation determines the method by which variables will be combined Arithmetic averaging is the most common and transparent method used to aggregate variables, and entails summing the product of each variable and its weight. Power averaging is a technique developed by Anand and Sen (HDR: 1997) in which variables are raised to a power alpha, summed with (usually equal) weights, and then the alphath root is taken. This technique has the effect of giving higher implicit weights to variables that are of higher levels, and is mathematically involved, which diminishes its transparency. A final technique for aggregation is multiplicative averaging which is used for calculation of conditional probabilities: for example, the risk of single parent poverty is estimated by the rate of divorce times the rate of single parent poverty.

The final step in forming a composite index is setting the weights within the aggregation scheme. We review four techniques for this process: societal determination from polls, surveys or focus groups, weights set by experts or policy makers, setting equal weights a priori before variables are chosen, and the use of Principal Component Analysis (PCA) to set weights.

We summarize the methodological techniques both in the typology of methodological choices (Exhibit 1) and a summary of the advantages and disadvantages associated with each choice (Exhibit 2). This leads us to three best practice suggestions for the construction of composite indices of well-being: use functional forms when the marginal changes in a variable are significantly different within the range of values taken on by the variable in the index; use a complementary approach to indices of social and economic well-being to highlight deprivation: both a conglomerative and a deprivational index; use Linear Scaling Technique (LST) to scale all variables before aggregating. Throughout the paper, we rely heavily on the work done by Sudhir Anand of the University of Oxford, and Amartya Sen, of the University of Cambridge and 1998 Nobel Laureate in Economics in assessing the UNDP's indices of human development.

Methodological Choices Encountered in the Construction of Composite Indices of Economic and Social Well-Being

MOTIVATION OF PAPER

The past decades have witnessed a dramatic increase in available statistical information measuring economic and social variables. This increase in data collection, in its international breadth and depth heightens the need for an interpretation and consolidation. Statistical indicators of social and economic well-being and of development have become a popular tool to consolidate and present data, and to indicate "progress" or "achievement" into a normative measure.

The use of statistical techniques to quantify complex concepts such as social health and human development tends to be bridged by economics and sociology, which seek to quantify measurable components of a society that determine social health or human development. These fields have motivated the use of numerical assessments of the progress or state of a society based on values on such categories. For example, in the 1980's, it may have been said that GDP per capita was the most important indicator of societal health. The 1990's have witnessed a move away from such reductionism, towards a growing body that acknowledges the multi-dimensional components of well-being. Moves towards a multi-dimensional component analysis of well-being and of a consolidation of these variables into fewer numerical components, requires the development of a statistically sound methodology of index construction.

The United Nations Development Programme (UNDP) has commissioned a series of reports on the methodological foundations of its indicators of human development. Two Economists, Sudhir Anand of University of Oxford and Amartya Sen, of Cambridge University and the 1998 Nobel Laureate in Economics have provided a comprehensive evaluation. While some of these contributions have been incorporated into the literature of social indices, several important contributions have not been discussed: namely, the importance of functional forms in index scaling and the use of complementary indices to mediate between a single composite index (as used in the Economic Freedom of the World, An Index of Social Health for Canada, the Index of Labour Market Well-Being, the Index of Economic Well-Being) and disaggregated indicators that map individual variables but make no effort to aggregate them (such as Vital Signs and The Social Health of a Nation). Indeed, a recent comprehensive review by Booysen is an excellent overview of well-known methodological techniques, yet fails to examine important methodology used by the UNDP in the Human Development Index and the Human Poverty Index. In addition, no major social and economic index outside of the UNDP incorporates the methodology developed by the UNDP, especially by Sudhir Anand and Amartya Sen.

The objective of this paper is to provide a comprehensive discussion of the choices involved in the construction of indices of economic and social well-being. We will outline the specific goals of an index of social and economic well-being: an ability to track changes, both in absolute and relative terms over time, and to have data in a format such that inter-country comparisons at a point in time and over time have meaning. That is, the goal of such an index would be to have both ordinal and cardinal significance over time and across countries.

Structure of Paper

We begin by discussing the method of choosing variables to measure social and economic well-being, and assess the importance of choosing variables with little correlation. We then discuss functional forms which capture the significance of marginal changes of variables at different levels including innovative approaches to scaling advocated by Anand and Sen (1994, 1997) in UNDP development indicators such as the logarithm. We then discuss the use of standardization of raw data in drawing conclusions over trends in time and the use of such data in inter-country comparisons. This leads into a discussion of the assignment of a number to a single variable, and the importance of numbers in their ability to represent change. We will discuss common approaches to standardization of variables: those that focus on emphasizing levels such as linear scaling transformations, models of data as Gaussian distributions, and ordinal analysis, and those that focus on emphasizing trends over time, such as normalization to the base year.

This is followed by a discussion of aggregation of standardized variables. We will outline different approaches to aggregation both from the perspective of analyzing the use of additive averaging, power averaging and multiplicative aggregation of variables as discussed in the literature. We will then discuss several methodologies employed to set weights of different variables. This will include a review of the technique of Principal Components Analysis as it relates to the inclusion of variables in an index, and the subjective interpretation of an aggregate. In particular, whether we view variables as the only available numbers to represent a larger picture of social or economic health, or as the object of study themselves. We will also assess the contribution individual variables to level changes in the index with different aggregation techniques. This will show interpretations of the relative importance of different variables, for example the ability of an increase in one variable to offset the decrease in another.
This will lead to a discussion of the complementary approach to index construction: the use of both conglomerative and deprivational indices together.

Our discussion will include an explanation of the Human Development Index (HDI) and the Human Poverty Index (HPI), which is as yet the most successful indices for the purpose of tracking temporal trends within a country, making cross-country comparisons and allowing for differential weighting of marginal increases at different levels of a variable. We also focus on the suggestion of Anand and Sen to adopt a complementary approach to indices—the use of both conglomerative and deprivational composites. This approach is taken by the UNDP: the HDI is conglomerative, measuring the total welfare of society and the HPI is deprivational, measuring the welfare of the worst-off.

We conclude by discussing the applications of the review of methodological techniques as they relate to the Index of Economic Well-Being and the Index of Labour Market Well-Being produced by the Center for the Study of Living Standards. Finally, we will make recommendations for best-practice approaches to the construction of composite indices of social and economic well-being. The paper builds on a recent and comprehensive review of methodological techniques by Booysen (2002).

VARIABLE CHOICE

We outline two main methodologies in variable choice: the selection of variables or variables up to the discretion of the index constructor in terms of the relevance of variables to what the index is seeking to measure, and the selection of variables by trying to minimize correlation between them. Because indices that seek to measure well-being can only rely on certain measurements from the real world, some interpretation of which variables to use is needed.

The first procedure is self-explanatory: variables are included in an index if they are relevant to concept that is being measured, for example, if inequality is being measured, the poverty gap and Gini coefficient would be included.

The second methodology in variable choice is to examine the correlation between variables in deciding which to include. Conceptually, two variables are correlated if they increase and decrease together. For example, if a rise in unemployment causes a rise in crime, crime and unemployment are correlated. In an extreme case, two variables might measure exactly the same quantity. For example, if there were no savings, and no social welfare programs, poverty rates and unemployment rates might be close to perfectly correlated. In this case, inclusion of both variables would be redundant.

The correlation is a statistic that measures the degree to which variables are correlated and ranges from 0 to 1. It may be useful for index constructors to calculate the correlation between all of the variables being used in the potential index and use this to assess the redundancy of variables. Of course, the level of redundancy that is acceptable is a subjective decision. For a more extensive treatment of the issue of correlation between variables in an index of well-being, see Srinivasan, 1994, Chakravorty, 2001, Lai 2000.

FUNCTIONAL FORMS

Significance of marginal changes in a variable

When interpreting the level of a variable, two issues arise: first, are absolute values of a variable proportional in importance for overall-well being; and second, are changes in the value of a variable of equal importance at various levels of the variable. The response to these questions leads us to consider functional forms: linear and non-linear. Functional forms represent the way changes in a variable are valued at different levels. If changes are valued in the same way, regardless of level, then the functional form should be linear. If changes are more significant at lower levels of the variable, the functional form should be concave down, such as the log or the nth root. If changes are more significant at higher levels of the variable, the functional form should be concave up, such as an exponential or power[1]. Both the functional

[1] The standard choice is for log as the concave down function and power as the concave up function.

form that is concave up and the functional form that is concave down are non-linear by definition.

In a recent paper surveying methodology, Booysen (2002) focuses the discussion on the statistical properties of different scaling techniques in the service of aggregation, and standardization, an underlying question remains unanswered: how are we to conceive of the importance of the numerical value of an individual variable? The significance of an absolute change at low level (say unemployment increase from 1 to 7 per cent) may be different than the same absolute change at a high level (say unemployment increase from 7 to 13 per cent).

The first step in index construction is to identify a generalized method of interpretation of raw data. The most common method for interpreting a value is that absolute changes in the variable become the most significant statistic: in other words, a change of $1000 in per capita GDP beginning at a high value (eg. from $20,000 to $21,000) is just as significant as a change of $1000 beginning at a low value (eg. from $5000 to $6000). If no functional form is applied to a variable and changes in level become the statistic used for an index, these marginal changes are valued in the same way, regardless of level.

Because of the importance of level comparisons in social indices, it is advantageous to apply functional forms to variables so that the marginal changes associated to the value after a functional form has been applied are consistent with the value of a marginal change in society. Sinden writes, "Constant marginal utility for increases in any variables is a highly unlikely phenomenon". Utility theorists, Anderson, Hardaker, and Dillon, argue that applying a functional form with decreasing marginal returns has more basis than a linear one (Sinden:1984 410).

Variables that are commonly taken into account in indices of social and economic well-being, such as per capita GDP, measures of unemployment, poverty gaps and rates, measures of inequality such as ratios of high and low incomes, and environmental depletion, are commonly thought to have significance of marginal changes that varies over the range of the observed values of the variable.

Anand and Sen (1997), state that, in measures of poverty deprivation " the relative impact of the deprivation .. would increase as the level of deprivation becomes sharper". According to this motivation, the UNDP develops measures of deprivation and inequality that more heavily penalize countries with higher indicators of deprivation in absolute value terms. For example, a decrease of 5 years of life expectancy from a base level of 40 is more heavily penalized than the same decrease beginning at a level of 80.

The UNDP approach: functional forms

The Human Development Index (HDI) is an excellent example of an index that uses a non-linear functional form in the construction of than index of well-being, and of the methodological implementation of this principle of diminishing returns in some measures of well-being, or, equivalently, increasing returns in measures of deprivation. The HDI is comprised of three components: per capita GDP, education achievement and life expectancy. In the HDI, the GDP index is calculated by taking the log of the GDP values per capita.

The motivation for this approach comes from Anand and Sen (2000), in which they survey the meaning of the income component of the HDI and its interpretation by the UNDP over the past 10 years. Since the first human development report appeared in 1990, the rationale for the HDI's treatment of the income component has been that it should reflect the fact that "people do not need excessive financial resources to ensure a decent living. This aspect was taken into account by using the logarithm of real GDP per capita for the income indicator" (HDR 1990 in Anand and Sen 2000), and the maximum and minimum values for income indicator were Zaire's per capita GDP and the average of the poverty line of 9 industrialized countries. In this original methodology, any country with per capita GDP above the poverty line of the average of 9 industrialized countries valued the same maximum value.

Anand and Sen argue that the goal of the 1991 HDR was " to introduce a variable elasticity valuation function that is both concave throughout the income range and for which the elasticity of marginal valuation increases with income" (1991: 93). In response, they analyze the elasticity marginal valuation function of W(y), where W(y) is the functional transformation of the original value of GDP. They develop a class of functions that can be applied to the per capita GDP to achieve a variable elasticity of marginal valuation that increases with income. In response to the 2000 report, the UNDP adopts a valuation function W(y)=log(y) which has the property of unit elasticity of marginal valuation (1991: 92). The 2000 UNDP states that this choice of W(y) "does not discount as severely as the formula used earlier. Second, it depicts income, not just the income above a certain level. Third as the figure asymptote starts quite late, so middle-income countries are not penalized… [except that per capita income above $40,000 is not counted]" (2000 HDR).

It is not clear that the analysis done by Anand and Sen can be applied to the case of highly developed economies, where the W(y) may be effectively linear in the range considered. For example, in the case of the GDP per capita, since the HDI incorporates a large range of income levels, it makes sense to take the log of the per capita GDP. However, the log of per capita income in OECD countries is essentially linear. The framework of Anand and Sen's analysis is of interest; it suggests the idea that other variables, whose range is larger in developed economies, might be subjected to a valuation function, such as unemployment or literacy level, and percentage of people living below the poverty line.

STANDARDIZATION OF VARIABLES

Scaling

Once variables and functional forms for the variables are chosen for an index, an essential question that underlies discussions of index methodology is should a single variable be scaled, and if so, what is the meaning or interpretation of a scaled variable. The essential reason why it may be necessary to scale variables is that raw data have significantly different ranges. In such cases, without scaling, composite indices will be biased towards variables with high ranges and meaningful changes in a value may insignificantly affect the composite index. Further, the unscaled aggregation of values is an implicit weighting scheme. For example, the UNDP's Human Poverty Index for Developed countries (HPI-2) aggregates four unscaled variables, among which are the long term unemployment rate and the percent of people lacking functional literacy skills. The range of values of percentage of people lacking functional literacy skills is three times the range of values of long term unemployment (UNDP: 2002). Since the variables are aggregated without scaling, there are higher implicit weights for overall change in ithe index composite put on the percentage of people lacking functional literacy skills.

Booysen (2002: 123), in a recent survey of methodological techniques, says that the "aim [of scaling variables] is to point out the relation among certain objects, how far apart they are and in what direction they lie relative to each other". Booysen outlines four possibilities for treatment of the scaling issue: no scaling, the use of normalized variables so that their mean is 0 and their standard deviation is 1, the use of ordinal response scales, and conventional linear scaling transformation (LST). We differentiate between standardization with an emphasis on transforming variables in order to standardize their range or variance and standardization of the base year level which emphasizes percentage change. The following classifications of methods to standardize variables are used: 1) no standardization, 2) normalization, 3) Z-Score or Gaussian normalization, 4) linear scaling, where ordinal ranking and LST are subsumed in the category of linear scaling. Note that LST scales variables to a common range, the Gaussian normalization scales variables to a common mean and standard deviation (0 and 1 respectively), and normalization scales variables to a common base year level.

Directionality Issue

A primary motivation for the standardization of variables is the fact that increases in some variables, such as literacy, correspond to increases in overall well-being, whereas increases in other variables, such as unemployment, correspond to decreases in overall well-being. We call this the directionality issue. We want to standardize variables so that an increase in the standardized score corresponds to increase in overall well-being. The procedures of Gausian normalization which produces a Z-score as the standardized variable and linear scaling which produces a scaled variable as the standardized variable, both provide methodologically consistent ways to standardize variables so that their increases correspond to increases in well-being. The technique of normalization to the base year is also able to deal with the directionality problem, but has other shortcomings. Without any standardization, the directionality issue is not resolved.

No Standardization

The first method, no standardization, would involve an aggregation of original data before being scaled. Booysen points out that this may be a good technique if all of the variables are percents or ratios. This technique is used by the UNDP in their Human Poverty Index (HPI-1, HPI-2), where original percentages are not standardized, and are aggregated into a composite index. However, if variables are not already in percentage or standardized terms, aggregating variables without standardization will cause the index to be dominated by implicit weights coming from the units and range used to measure variables. As discussed above, even in the UNDP's HPI, which aggregate percent values, no standardization results in implicit weighting.

Normalization to base year

The technique of normalization is used in the Index of Economic Well-Being. Each variable is normalized to the first year where data are available, and these normalized values are aggregated. This technique is essentially one aggregating percent changes over time in each variable. The advantages are the percent changes over time are highlighted, which is valuable for tracking temporal trends. It is also possible to deal with the directionality issue by taking the reciprocal of standardized variables whose decrease corresponds to increases in well-being, and aggregating these variables. Using this procedure, if the unemployment rate doubles from the base year then the standardized value will half, and this is the value that will be aggregated.

The disadvantage is that variables with low bases compared to the range of values can skew the index and cause small absolute changes in this variable to overwhelmingly effect the composite. For example, if the unemployment figures range from .5% to 15%, a change from 0.5% to 15% is a 30 fold increase, and a change from 0.5% to 5% will be a ten fold increase. However, in a different range of data, say between 10.5% and 15%, the same absolute change, of 4.5% from 10.5% to 15% is less than a 1.5 fold increase.

The Z-Score

The Z-score is calculated subtracting the mean of a data set and then dividing by its standard deviation. The technique is based on the class of functions called Gaussian curves. A normal or Gaussian curve is defined by parameters of mean and standard deviation, and given by the formula $f(x)=\frac{1}{\sqrt{2\pi\sigma}} e^{(x-\mu)/\sigma}$ where μ is the mean and σ is the standard deviation. If the standard deviation (square root of the variance) is calculated for a set of variables with mean zero, and then all values are divided by the standard deviation, the resulting set of values will have standard deviation 1. The procedure of Gaussian normalization first subtracts the mean of the variables, so that the mean of the transformed variables is zero.

Any variable standardization procedure must deal with the fact that for some variables, increases correspond to increases in well-being, and for others, increases correspond to decreases in well-being. The Z-score technique deals with this issue by multiplying the scaled variable by negative 1 if increases correspond to decreased levels in well being. However, Gaussian normalization does not necessarily standardize all variables to a specific range. Because the procedure transforms each variable by subtracting the mean and dividing by the standard deviation of a data set, depending on the data, it is likely that some data points will be outside of one standard deviation. The Z-score of a variable represents the number of standard deviations it is away from its mean, and therefore does not standardize variables to a common range.

For example, suppose that the unemployment data for 10 years and 10 countries is comprised of 48 values clustered near 5% unemployment, 48 values clustered near 8% unemployment and 1 value of 1% and one value of 15%. The mean unemployment rate is 6.53%. So after the mean is subtracted from the data (the first step in Gaussian normalization), the values are 48 at –1.53%, 48 at 1.47%, 1 value at 8.47% and 1 value at –5.53%. The standard deviation is thus $(1/100)*(48*(-1.53)^2+48*1.47^2+8.47^2+(-5.53)^2)^{1/2}=1.36$. Thus, the Z-scores for these values are 48 at –1.125, 48 at –1.08, and 1 each at 6.23 and -4.067. Notice that these values do not represent a uniform range, and that the range is determined by the initial distribution of values.

Gaussian normalization is used by the World Economic Forum in its 1996 Global Competitiveness Report (WEF: 1996), and GeoAccess Division of Natural Resources Canada in its Atlas of Canada Quality Of Life Mapping Project (Morton: 2002).

Ordinal response

Ordinal response is the technique where experts or evaluators interpret variables and classify them according to ordinal scales, usually between 1 and 5 or 10. Each factor in this index is valued to be in such a range by the people responsible for the construction of the index. Higher values correspond to more economic and social well-being and lower values to lower economic and social well-being. An example of ordinal response scales is found in the World Competitiveness Report 2002, where surveys ask for ordinal rankings of variables, and responses are averaged.

Linear Scaling Technique

Linear Scaling Technique (LST) is a technique used to standardize the range of a variable. To do this, an estimate is made for the high and low values which represent the possible range of a variable for all time periods and for all countries, and denoted Min and Max, respectively. The data is then scaled according to these values. If a variable increase corresponds to an increase in overall welfare, the variable, VALUE, is scaled according to the formula

1) $\frac{\text{Value-Min}}{\text{Max-Min}}$

In this case, we see that increases in the VALUE correspond to increases in scaled VALUE. Notice that if the Min is equal to zero, the formula above reduces to VALUE/Max.

If, in contrast, an increase in VALUE corresponds to decrease in overall welfare, the VALUE is scaled according to the complementary formula,

2) $\frac{\text{Max-Value}}{\text{Max-Min}}$

In this case, we see that increases in the VALUE correspond to decreases in the scaled VALUE. In both cases, the range of values is 0-1, and 0 corresponds to the lowest level of welfare, and 1 corresponds to the highest. Note that this formula reduces to (Max-Value)/Max when Min is set to 0. This technique is used to scale all variables in many indices, including the following: the Human Development Index produced by the UNDP, An Index of Social Health by Human Resources Development Canada (HRDC), the Index of Economic Freedom by the Heritage Institute and Economic Freedom produced by the Cato Institute.

AGGREGATION OPERATION

In addition to selecting variables, their functional forms, and a method of uniform standardization, an operation to combine components must be chosen. In order to aggregate consistently, we must standardize all variables contributing to the aggregate in the same way. A standard approach to aggregation is the addition of all components to form the composite index. Because standard averaging is straightforward, we will concentrate on multiplicative aggregation and the use of power-averaging developed by the UNDP.

Multiplicative Aggregation

Developers of an index of social or economic well-being may want to include a variable quantity such as risk, that is a conditional probability and cannot be directly measured by a single variable alone. For example, the Index of Economic Market Well-Being seeks to measure the risk of single parent poverty. The only available variables are poverty incidence of single parent families and the rate of divorce. In order to find the rate of single parent poverty, we need to consider conditional probabilities. That is, the probability of being a single parent in poverty is modeled as the probability of being in poverty if you are divorced, times the probability of being divorced[2]. For this reason, the index measures the rate single parent poverty as the product of the rate of divorce times the rate of poverty among single parents.

This procedure has the advantage of more accurately quantifying the risk of poverty in society, but has some methodological problems. For one thing, if variables are scaled with LST before they are multiplied, the overall risk will not be scaled to the same range as the original variables. For example, the maximum and minimum levels of two variables may not ever be present in one single measurement, causing the multiplicative range to overestimate the actual range of the product of the two variables. Suppose that in 1980 the risk of divorce was .4 and the risk of single parent poverty was .8 and that in 2000, the risk of divorce was .8 and the risk of single parent poverty was .4. Then the total range coming from scaling variables before multiplying them would be .16 to .64, but in reality, both values are at .32. This problem could be overcome if conditional probabilities were first multiplied and then standardized.

Additive Averaging

Additive averaging is a technique of aggregating variables that gives explicit weights to each variable and sums the product each variable and its weight. The sum of all of the weights must be 1. This technique is advantageous because of its methodological transparency. For example, if four variables A, B, C, D are average with equal weights, the average is simply 0.25A+0.25B+0.25C+0.25D. And example of different weights would be 0.4A+0.4B+0.1C+0.1D, and any choice of weights which together add to 1 could be used. Because of its transparency and frequent usage, we do not discuss the method further here.

---

[2] This model assumes that single parents were once married.

Power Averaging

The UNDP's Human Poverty Index 1, and 2 (HPI-1, HPI-2) use a different approach, that of power-averaging, which will be explained below. Instead of simply averaging the components that comprise the index: P_1=the probability of not surviving to age 60 (times 100), P_2 = adults lacking functional literacy skills, P_3=population below the poverty line and P_4=long term unemployment rate (lasting 12 or more months), power averaging is used. The method employed is the power averaging method, discussed extensively in the 1997 HDR (UNDP: 1997), in which variables are aggregated according to $(1/3(x^{\alpha}+y^{\alpha}+z^{\alpha}))^{1/\alpha}$ [3]. This means that first, each variable is raised to the alpha power, then the terms are added and multiplied by 1/3 and then the alphath root is taken. The motivation for this methodology is described in the technical notes to the 2001 HDR, "the value of 3 is used to give additional but not overwhelming weight to areas of more acute deprivation." The same methodology is applied in the Index of Multiple Deprivation produced by the Social Policy and Development Centre in Pakistan (2002:82). The goal of power averaging is to weight a country's deprivation score more heavily in the area of most significant deprivation.

One may wonder why the 1/ α root is taken after the variables are raised to the α power and then averaged. This methodology is motivated by the so called L-p norms in classical mathematical analysis. We will not concern ourselves with the technical details, although the subject is an area of extensive study in mathematics. One reason to take the alphath root after the variables are raised to the α power and averaged is that if all variables which are averaged have the same value, then power averaging will give exactly the same result as simple additive averaging. Power averaging of 4 variables with the same value gives a formula that looks like $(1/4\ (x^3+x^3+x^3+x^3))^{1/3}=(1/4\ (4x^3))^{1/3}= x$, and simple additive averaging gives a formula that looks like $(1/4*(x+x+x+x))=x$.

Anand and Sen[4] (HDR: 1997) give a rigorous analysis of the technique of power-averaging without specifying alpha. Their discussion is structured by qualitatively proscribing several properties that a human poverty index should have, and showing that the mathematical formula is consistent with this. They define a function P(α) which is H above, but varies with α. P(α) is interpreted "as the degree of overall poverty that is equivalent to having a headcount

---

[3] See 1997 Human Development Report for a complete mathematical characterization.

[4] For calculating the HPI (and gender development index, GDI), α =3 is chosen because it mediates between not completely privileging the effect of an increase in a higher variable (low α) and having the property that as poverty indices go up, their relative significance also goes up. However, the authors acknowledge that the choice of any α is inherently arbitrary.

The effect of different values of alpha is explored in the technical notes to the 1997 index. The interpretation of change in variables X,Y,Z and their substitutability in the formula, $1/3(X^{\alpha} + Y^{\alpha} + Z^{\alpha})$^ $^{1/\alpha}$, changes as alpha changes. First, note that when alpha is 1, this is simply the arithmetic average of X, Y, Z. In this case, changes in X,Y,Z are indistinguishable. Anand and Sen introduce the proposition 9 (1997: 121): the elasticity of substitution \sigma between any two subindices of P(α), that is between any two of X,Y,Z is constant and equal to 1/(α -1). IN other words, when α is 1, there is perfect substitutability, and when alpha is infinity, there is no substitutability (1997: 121). The authors note that "the usual assumption that as the extent of deprivation in any dimension increases, the weight on further additions to deprivation in that dimension should also increase. For this we need α >1. The increase in X compared to an increase in Y is $(X/Y)^{\alpha -1}$" (1997: 121).

ration of $P(\alpha)$ in" each of the three dimensions. The following property is desired in an aggregate poverty index with distinct subindices: 1) the aggregate index should increase with increases in each factor. The increase of the index should be at an increasing rate as each component grows, that is, it should be convex with respect to each of the factors. In other words, is should also decrease at a diminishing rate with each factor. Another concept to consider is the disaggregation of $P(\alpha)$, which turns out is not strictly decomposable. In other words, $P(\alpha)$ of groups of disaggregated variables is smaller than the weighted mean of the $P(\alpha)$ on each subindex. The authors go on to calculate the elasticity of substitution between P_i and P_j, factors composing $P(\alpha)$, and that a relative unit increase in P_1 compared to P_2 is given by $(P_1/P_2)^{\alpha-1}$.

For calculating the HPI (and gender development index, GDI), alpha=3 is chosen because it mediates between not completely privileging the effect of an increase in a higher variable (low alpha) and having the property that as poverty indices go up, their relative significance also goes up. However, the authors acknowledge that the choice of any alpha is inherently arbitrary.

Example of Power-Averaging in the Human Poverty Index

The Human Poverty Index (HPI-1 and HPI-2) use alpha averaging with alpha taken to be 3. In this section, we give a comparison of the rank order and scores of the formula used by the UNDP (alpha=3 averaging) and the formula of simple averages.

A particular example will be the sample calculation for Australia in the 2001 HPI-2, 4 aggregated probabilities described above, P_1= 9.1 per cent, P_2=17 per cent, P_3=2.1 per cent, P_4=14.3 per cent, are calculated in the following way: HPI-2= $(¼*(9.1^3+17.0^3+2.1^3+14.3^3)^{1/3}$=12.9. If we were to do simple aggregation on these variables, we would calculate ¼*(9.1+17.0+2.1+14.3)=10.6. If we use the United States as an example, we see a another difference: The HPI-2 (2002) = $(1/4* (15.8^3+12.8^3+20.7^3+.2^3)^{1/3})$=15.8, whereas the average of these values is = ¼*(15.8+12.8+20.7+.2)=12.3.

Below, we give a table of a comparison between the method of calculation for the HPI-2 and a simple average of the variables, and compare their rank. (Data from 2001[5]):

[5] The Probability at birth of not surviving to age 60 is given as per cent of cohort, People lacking functional literacy skills is given as per cent of people between the ages of 16 and 65, Long-term unemployment is as a per cent of labour force. Hence, all variables have a potential range between 0 and 100 per cent.

Table 1: Comparison of power averaging and standard averaging with 2001 HPI-2 data

	Probability at birth of not surviving to age 60 (%)	People lacking functional literacy skills (%)	Long-term unemployment (%)	Percent of people below 50per cent of median income	HPI-2 Calculation (3rd power averaging) [6]	HPI-2 Rank	Arithmetic Average	Rank
Norway	9.1	8.5	0.2	6.9	7.5	2	6.2	1
Sweden	8	7.5	2.8	6.6	6.8	1	6.2	2
Canada	9.5	16.6	0.9	11.9	12.1	7	9.7	7
Belgium	10.5	18.4	5.5	5.2	12.4	8	9.9	8
Australia	9.1	17	2.1	14.3	12.9	9	10.6	9
US	12.8	20.7	0.3	16.9	15.8	12	12.7	12
Netherlands	9.2	10.5	1.4	8.1	8.5	3	7.3	3
Finland	11.3	10.4	3	5.2	8.8	4	7.5	4
UK	9.9	21.8	1.8	13.4	15.1	10	11.7	10
Denmark	12	9.6	4.5	7.2	9.2	5	8.3	5
Germany	10.6	14.4	5.6	7.5	10.6	6	9.5	6
Ireland	10.4	22.6	5.6	11.1	15.3	11	12.4	11

Data from UNDP's Human Development Report 2001 HPI-2 2001.

Notice that the only difference in relative rankings is between Sweden and Norway. Sweden is ranked first by the HPI-2 because its measures of deprivation, while having a slightly higher mean than Norway (see table), have much smaller average of their third powers, because they are more closely distributed to the mean. Another observation is that the HPI-2 index widens the range of the total index compared to a simple average of the values. In the above example, the HPI-2 has a range between 6.77 and 15.8, whereas the simple average is between 6 and 12.6. With more varied data, this trend might be more pronounced. But it is worth noting that the ranks of the countries are almost the same with either of the formulas, which leads us to consider whether the methodological transparency lost by using alpha=3 outweighs its technical sophistication.

We now give an example of the HPI-1 with selected countries in table format.

---

[6] If X, Y, Z, W denote the variables in the columns, the HPI-2 is calculated as $(1/4*(X^3+Y^3+Z^3+W^3))^{1/3}$

Table 2: Comparison of power averaging and standard averaging with 2001 HPI-2 data

	Probability at birth of not surviving to age 40 (%)	Adult illiteracy rate (%)	Population not using improved water sources (%)	Underweight children under age 5 (%)	HPI-1 (3rd power average)	HPI-1 Rank	Average	Average Rank
Pakistan	20.1	56.8	12.0	38.0	41.0	7	34.0	3
Sudan	27.3	42.2	25.0	17.0	32.7	1	30.2	2
Togo	34.1	42.9	46.0	25.0	37.9	5	37.5	8
Nepal	22.5	58.2	19.0	47.0	43.4	10	37.9	9
Yemen	20.0	53.7	31.0	46.0	41.9	8	37.4	7
Bangladesh	21.4	58.7	3.0	48.0	42.4	9	35.2	5
Haiti	31.6	60.2	54.0	28.0	47.4	11	44.3	11
Madagascar	31.6	33.5	53.0	33.0	36.7	4	36.0	6
Nigeria	33.7	36.1	43.0	27.0	35.0	3	34.9	4
Djibouti	42.3	35.4	0.0	18.0	34.3	2	28.9	1
Uganda	48.4	32.9	50.0	26.0	40.8	6	39.8	10

Data from UNDP's Human Development Report 2002 HPI-1 2002, see technical notes for methodology.

This example better illustrates differences in ranking that result from power-averaging. The rankings of Pakistan, Bangladesh and Uganda change 4 ranking places between the 3rd power average and standard averaging. This is due to the distribution of deprivation in the variables. In Pakistan and Bangladesh, the low level of people using unimproved water sources causes the total average of the components in the HPI-1 to be small. However, because both countries have high levels of other variables measuring deprivation, the 3rd power averaging of the components lowers their ranking. The opposite is the case for Uganda, which has its deprivation more evenly distributed.

EXPLICIT WEIGHTING OF VARIABLES

In the above methods of additive averaging and power-averaging, explicit weights must be chosen. In the discussion of power-averaging above, the choice of equal weights was made implicitly by dividing by the total number of variables. However, other weights can also be chosen. We now discuss a method for determining these explicit weights in a composite. As discussed by Booysen (2002), the following techniques to set explicit weights in aggregation are the most widely accepted and used[7]:

- Expert weighting set by specialists, or societal determination
- Principal Component Analysis.

[7] Note that implicit weights are set in scaling.

- Explicitly set weights by another mechanism, such as equal weighting

Expert weightings are typically set by a group of specialists who set weights for each component. The values determined by specialists are then averaged. Weights are sometimes set by policy makers or social surveys about how meaningful or important individual variables are to people. Other explicit weighting systems exist, such as the explicit equal weighting of all variables in an aggregation.

To begin the discussion, we will consider two different interpretations of variables in the construction of aggregate indices. The first is the interpretation that view variables as ends in themselves, or that an aggregation of variables measures exactly what an index is interested in assessing. An example of this might be an aggregation of educational statistics, where literacy rate, enrollment and graduation rate, say, are exactly the quantities in which one is interested. This interpretation is in contrast to the idea that variables serve as indicators of a concept that we cannot measure with current tools. In this case, variables are variables. An example of this is the idea that a variety of diverse variables are all indicators of well-being, but no set of variables can perfectly capture the concept of social or economic well-being.

The literature on setting weights to aggregate variables, and on assessing the value of a weighting scheme tends to center on the process of setting weights as opposed to the effect of setting weights on the aggregate value. This aspect of interpretation is also meaningful: that if a survey has 4 components, each scaled with LST, the 1st weighted with 10 per cent, the 2nd weighted with 20 per cent, the 3rd weighted with 30 per cent and the 4th weighted with 40 per cent, these weights imply that if the 1st scaled value increase by 0.4 points, and the 4th decreases by 0.1 point, the aggregate will remain the same. In other words, in a single index, implicitly, weights themselves determine the contribution of a variable to well-being. In this way, we notice that a linear aggregation of n variables results in n equations equating changes in scaled values, where if the weighting scheme is $a_1 * x_1 + a_2 * x_2 + \dots a_n * x_n$, then we have implicitly the equations a change of 1 point in $x_1$is equivalent to a change of a_n/a_1 points in x_n, a_{n-1}/a_1 points in x_{n-1}, and so on. This gives n-1 equations relating equivalences of changes in the levels of the n-1 variables, scaled by LST.

Another artifact of these equations is that they imply there is transitivity among the exchangeability among variables, in the same way as utility valuations imply transitivity among preferences. For example, suppose that we take the variables enrollment in secondary school, enrollment in tertiary school, infant mortality rate and life expectancy. It is imaginable that we might be able to relate the variables of secondary and tertiary enrollment to each other and infant mortality and life expectancy to each other with linear equations insofar as we could conceive of a relative weight between changes in each pair. However, it is much harder to relate the importance of changes in life expectancy to the importance of changes in secondary school enrollment.

Principal Component Analysis (PCA)

It is perhaps this difficulty that has led some to use the technique of principal component analysis to weight variables in a composite index. The goal of principal component analysis is essentially to uncover variations in a data set. Principal component analysis can be used to describe the variation of a data set using a smaller number of dimensions than number of variables of the original data. The weights of the components in the first principal component, which we call the principal component, are assigned to maximize the variation in the linear combination of original variables, or (equivalently) to maximize the sum of the squared correlations[8] of the principal component with the original variable. Another way to think about this is that the first principal component is represented by the line in the original space of variables that minimizes the sum of the squared distances between it and the original data points.

PCA can be used to set weights in a set of data by using the coefficients of the first principal component as weights. In fact, PCA is a linear algebraic technique which generates weights of different variables by giving them exactly the components of the first eigenvector of the covaraince matrix. The weights assigned by PCA to various categories are therefore the weights that, give the maximum variation in the aggregated index values, over all possible choices of weights.

Although correlation PCA has some mathematical sophistication, its use in weighting components of social indices is dubious. For example, it may lead to variables in a data set which have little variation being assigned small weights. However, the index constructor chose variables that were deemed to be As has been pointed out, PCA removes control over the weighting of individual variables from the people who construct the index (Booysen references Ginsberg 1986), and gives a index a false aura of mathematical objectivity, one that may never be achieved a social index.

On the other hand, expert weighting, or weighting done by policy makers or public opinion polls suffer from disadvantages as well. As discussed before, an assignment of weights that implies the knowledge of relative valuation of variables is difficult, especially when different facets of social and economic well-being are aggregated. Policy makers or government officials will most likely be unable to represent their constituents relative valuations of components in a social index. And constituents may be unlikely to be able to compare such disparate aspects of well-being; indeed, this may be impossible, as peoples' preferences can be non-transitive, especially with high number of variables. For this reason, it may be wise to abandon the notion that there exists a set of weights capable of perfectly capturing the relative contribution of variables to overall well-being.

Equal Weighting

In light of the difficulties surrounding the explicit third party determination of weights, as well as the lack of interpretive meaning for PCA, we should consider turning our attention to the idea that all variables should a priori be weighted equally. Although equal weighting is certainly an explicit weighting scheme, the a priori decision to adopt the technique of equal weighting for methodological purposes makes the choice of weights less subjective. A motivation for this approach is that it is objective in the sense that if adopted as a common technique of index aggregation, the subjective component of construction of indices would lie exclusively in the

[8] The correlation matrix is the covariance matrix of variables which are scaled in order to have unit variance

choice of variables. There is an advantage of this approach: namely, that a debate over the inclusion of variables, that is, which variables are important, can be conducted on a more basic level than a discussion that centers around the choice of numerical weights.

Another strength of this approach is that if variables are chosen as variables for something that cannot be perfectly quantified such as inequality, from the perspective of the social indicator constructor, the variables chosen as variables for a category of measurement should form a collection of multidimensional variables that is a sampling of indicators that may represent the category. Since the variables are variables and not measurements in themselves, it is more consistent to treat them as statistical objects that are not subject to further subjective numerical interpretation. As discussed earlier, involved statistical techniques such as PCA or factor analysis do not always make sense to apply to social indices due to the complex nature of social and economic phenomena. As a result, the case for uniformly aggregated variables, that is, a priori equal weights, is strengthened.

SINGLE VERSUS COMPLEMENTARY APPROACH TO COMPOSITE INDICES

The UNDP uses the literacy variable (percent of population unable to read) in two different indices that are used to measure development, both of which are applied to developed countries: it appears both in the HDI and in HPI-2[9]. In the HDI, it is aggregated after it has been linearly scaled, and in the HPI-2, it is aggregated after it has been raised to the power alpha=3, as discussed above. The HDI and HPI-2 are what we will call complementary indices.

The example of the HDI

The HDI represents what is called a "conglomerative" approach to measures of social welfare, that is, a measure that evaluates the average welfare of a society. In the conglomerative approach, it may be possible for a great increase in the welfare of the best off to offset a decrease in the welfare of the worst off. The HPI-2 reflects a "deprivational perspective" in which the poverty of the worst-off members of society is underlined, and cannot be offset by an increase in the welfare of the best-off. The motivation for the complementary approach is outlined by Anand and Sen, in which they argue that the conglomerative and deprivational perspectives are not substitutes of each other. "We need both, for an adequate understanding of the process of development. The plurality of our concerns and commitment forces us take an interest in each" (1997: 2).

The above passage provides the motivation for the two essentially different indices of development used by the UNDP: first, the HDI, which is comglomerative, and in which increases in GDP possibly representing unequal gains per capita, can, albeit with diminishing returns, offset increases in illiteracy; second, the HPI, a derivational measure where increases in the welfare of the best-off do not offset decreases in the welfare of the worst-off[10]. It is

[9] The Human Poverty Index, HPI, is calculated differently for countries that are considered developing and countries considered to be developed. The former index is called the HPI-1 and the latter is called the HPI-2.

[10] The 2002 HPI-1 (for developing countries) is comprised of three equally weighted components: probability at birth of not surviving to age 40 (times 100), adult illiteracy rate, unweighted average of population not using

important to point out that variables to measure the welfare of the worst-off may not be available in a standardized format across countries. For example, even measures such as life expectancy at birth may not measure the welfare of the worst-off economically in a country that is ravaged by AIDs. Variables that represent deprivation in the HPI-2, such as illiteracy rate and mortality rate are good but not perfect indicators the welfare of the worst off. The percentage of people below 50% of the median income, a variable also used in the HPI-2, is a more consistent measure of the welfare of the worst off. The HPI-1 for developing countries, which is the companion to the HPI-2 for developed countries, uses two variables that more consistently represent the welfare of the worst off: percentage of population not using improved water facilities and percentage of children under 5 who are underweight. It also includes the probability at birth of not surviving until age 40 and adult illiteracy rate. See Table 1.

Ethical motivation for complementary indices

The analysis leading to complementary indices: deprivational and conglomerative is founded on extensive work in ethics and economics, in particular by Rawls and Sen. The methodology leads us to consider two ideas for modification of indices of social and economic well-being.

The first is that indices of social and economic well-being themselves should reflect the conglomerative and disaggregated approaches. According to this methodology, deprivational and conglomerative indices should be constructed separately side-by-side along the lines of the UNDP indicators. For example, the Index of Economic Well-Being, which is currently a conglomerative index, could be supplemented with a deprivational measure including variables of income inequality, risk from poverty in old age, risk of poverty, illness and unemployment. For indices such as the Index of Economic Well-Being, the interpretation of 'deprivational" may be different than in the HPI. Here it may be better interpreted to be social costs or risks, and include a subset of the variables already used in the conglomerative index that represent measures of deprivation: for example, inequality, environmental degradation, and foreign debt.

In addition to the ethical dimensions that motivate the complementary index approach, we see a statistical motivation. General aggregations of variables using a conglomerative strategy do not reflect the variability of the components of the index. In other words, if 4 values are averaged, distributions of 21, 1, 1, 1 and 6, 6, 6, 6 equally influence the index, although the values themselves are likely to represent vastly different social conditions. The Index of Economic Well-Being serves as a good example of this, where the four categories are consumption flows, wealth stocks, equality and economic security. The introduction of a deprivational index would highlight a difference between these distributions.

Of course, this approach is not able to differentiate all distributions, and we could imagine two quite different distributions that value similarly on both conglomerative and deprivational measures. However, if the goals of a social index are to both evaluate the total

---

improved water sources and underweight children under age 5. The HPI-2 is comprised of 4 equally weighted components: probability at birth of not surviving to age 60 (times 100), adults lacking functional literacy skill, population below income poverty line, long-term unemployment rate.

well-being of society and to assess the conditions of deprivation and risk within it as (REFS) suggest, the complementary approach may be valuable.

APPLICATIONS TO INDICES OF SOCIAL AND ECONOMIC WELL-BEING

Summary of aggregation techniques and their discussion in the literature

The above discussion of UNDP methodology of index construction highlights two techniques that are used in the HDI and HPI-2 but are not found in other major social indices[11]: the use of functional forms before applying the Linear scaling transformation (eg. the logarithm or the power function), and the introduction of the distinction between and incommensurate nature of conglomerative (HDI) and deprivational indices (HPI).

Little attention is paid to these techniques in the literature. Instead, criticism of the HDI focuses on a discussion of the appropriateness of variable choices, and the chosen weighting scheme. In 1992, Tatlidil used Principal Component Analysis to argue for equal weights on the three measures (Tatlidil 1992). Another discussion centers on whether the three component variables are correlated and how their weights should be adjusted. For example, Allen Kelly (1991) suggests that the weights should be set according to a “meta production function”. Because such is not available, he advocates the exploration of variable correlations. Srinivasan has criticized the index because, even if PCA is used, the high correlation of variables, the linear combination given by PCA “says nothing about what aspects of development are being portrayed by the combination” (Srinivasan 1994). McGillivray and White (1991) have found symmetry and lack of causality in the correlation of the variables.

Criticisms of the HDI’s weighting and measuring methodology are primarily structural critiques, such as the idea that all data should be first scaled to be spread around the same mean with the same variance so that the three factors can be represented as vectors in 3 dimensional space. Once this transformation is done, the vector distance between the country with the most desirable vector could be measured and give the values (Noorbakhsh AMHDI 1998). Other critiques focus on the correlations between the actual variables: in another paper, Noorbahakhsh claims that the HDI has the desired property that the components in the HDI are not highly correlated with each other and that the index is not highly correlated with any of its components (THDI: 1998).

Suggestions for changes in the Index of Economic Well-Being and the Index of Labour Market Well-Being

We now evaluate the applicability of a complementary approach for two indices constructed by the Centre for the Study of Living Standards: the Index of Economic Well-Being and the Index of Labour Market Well-Being. Both the Index of Economic Well-Being and the Index of Labour Market Well-Being are conglomerative indices. We notice that in the HPI (the deprivational index), variables are not scaled before they are alpha-averaged. This choice might

[11] An exception is Social Development in Pakistan 2001 (SPCD 2002).

have been made for methodological transparency. But it has the undesirable effect of using unscaled data, which suffers from the problem of implicit weighting due to measurement scale. To make the approach to the Index of Labour Market Well-Being and the Index of Economic Well-Being complementary, we now suggest a deprivational component to these indices with sample calculations which use scaled variables.

A complementary approach to Index of Economic Well-Being

The variables in the Index of Economic Well-Being are market consumption per capita, government spending per capita, variation in work hours, capital stock per capita, R&D per capita, natural resources per capita, human capital, net foreign debt per capita, social cost of environmental degradation, poverty intensity, Gini coefficient, risk from unemployment, risk from illness, risk from single parent poverty, risk from poverty in old age.

The deprivational index complementary to the conglomerative one already in existence can use variables already used in the conglomerative index but which specifically rep\flect the well-being of the worst-off. This suggestion is along the lines of the complementary approach used by the UNDP: the conglomerative measure of the HDI and the deprivational measure of the HPI. We suggest the deprivational aspects of this index be risk from poverty in old age, risk from single parent poverty, risk from illness, risk from unemployment, poverty intensity, and Gini coefficient, and social cost of environmental degradation.

A complementary approach to the Index of Labour Market Well-Being

The variables in the Index of Labour Market Well-Being are labour compensation per worker, labour compensation per hour, average educational attainment, hourly wage inequality, incidence of low wage unemployment, average of overall and long term unemployment rate , and a measure of the unemployment benefits rate.

We suggest the deprivational aspects of this index be hourly wage inequality, incidence of low wage unemployment, average of overall and long term unemployment rate, risk from workplace injury and death, risk from poverty in old age.

Below (see attached chart), we present a deprivational index for the Index of Labour Market Well-Being. We aggregate the index according to averaging its components, as well as with α =3 to compare. Notice that the choice of α =3 has a much more pronounced effect on the number for the US. While Switzerland's total conglomerative index value ranges from 0.5 to 0.6, the US' has a range near 0.4. The deprivational index, using the simple average makes this difference starker: Switzerland values about 0.8, whereas the US does not reach 0.5; with α =3, the difference is even sharper with the US not even achieving 0.3, where Switzerland maintains the same range.

CONCLUSIONS AND BEST PRACTICE

The main attention of the paper has been to provide a systematic interpretation of methodological choices presented in the construction of an index of well-being. These choices determined the structure of the paper, and are exemplified in Exhibit 1, a Typology of Methodological Choices Associated with Composite Index Construction. We have tried to provide an analysis of the rationale behind each methodological choice, as well as summary of the advantages and disadvantages of using each approach. This is summarized in Exhibit 2. To conclude the paper, we will review each of main choices encountered—single vs. complementary approach, variable selection, functional forms, standardization method, aggregation operation, weighting scheme—and suggest best practice procedures in some cases. We evaluate these choices for methodological transparency and technically consistent methodology.

- Best Practice: Use functional forms when the marginal changes in a variable are significantly different within the range of values taken on by the variable in the index.

The use of the functional form of log in the HDI, that is, scaling the log of the per capita GDP values is methodologically consistent with the idea that per capita income has sharply decreasing returns to scale. This is an important point made by the UNDP, and the use of the log functional form is a methodologically consistent way of representing this normative view. The use of the log functional form in the HDI is fairly transparent, and is the only non-trivial functional form used on the variables: the other data is simply scaled with LST. However, it is not clear whether the use of a combination of non-trivial power and log functional forms on many different variables in a single index would be able to maintain the same transparency. We suggest that applying non-trivial functional forms is best-practice when the marginal changes in a variable are significantly different within the range of values taken on by the variable in the index.

- Best Practice: Use a complementary approach to indices of social and economic well-being to highlight deprivation: both a conglomerative and a deprivational index.

Alpha-averaging in the HPI, a deprivational index, is used to give higher implicit weights to areas with higher levels of deprivation. The use of alpha greater than 1 in a deprivational index corresponds technical consistency. On the other hand, since the motivation for setting alpha greater than 1 requires involved mathematical derivation, alpha equally 1 gives rise to a more transparent methodology. To justify technical consistency, we would need a convincing difference in between rankings derived from the use of alpha =1 and alpha =3. Indeed, in the small ranges we are considering, where variables range from 0 to 1, the true difference between these two schemes may not be very significant. In Table 1, we saw that for the 2001 HPI-2, the difference between alpha=1 and 3 only changed one pair of rankings. In the 2002 HPI-2, the rankings for alpha=1 and 3 were identical. Of course, this is not a mathematically rigorous method of evaluating the choice of a value of alpha. But it gives convincing circumstantial evidence that the advantages from transparency of choosing alpha to be 1 may outweigh its methodological consistency of choosing it to be 3.

The use of a complementary approach, conglomerative and deprivational, however, does not suffer from the same loss of methodological transparency. In fact, the inclusion of a deprivational measure may help clarify the meaning of a conglomerative index and vice versa. The reason for this is that a deprivational measure can clarify whether a high level of well-being is concentrated in a well-off section of society, or whether a lower level is more evenly spread. Similarly, a conglomerative index can give more descriptive nature to the total well-being of society which may influence the potential for alleviating that deprivation. A complementary approach to index construction may be a powerful and universalizable tool for social index methodology.

- Best Practice: Use the Linear Scaling Technique to standardize all variables

By standardizing the range of variables, Linear Scaling Technique assigns the lowest implicit weights to variables of all procedures we considered. It also deals with the directionality issue, and provides a consistent way to aggregate variables some of whose increases correspond to increases in well-being, and some of whose decreases correspond to increases in well-being.

Table 3: A Deprivational Index for the Index of Labour Market Well-Being, Center for the Study of Living Standards, Comparison between Switzerland and the United States

Switzerland	Scaled index of 9th to 1st decile	Scaled incidence of low-wage unemployment	Scaled standardized unemployment rate	Scaled long-term unemployment rate	Scaled injury rate	Scaled fatality rate	Arithmetic average	Power average (alpha=3)
1980	0.27	0.43	0.02	0.01	0.16	0.18	0.18	0.26
1981	0.27	0.43	0.02	0.01	0.16	0.18	0.18	0.26
1982	0.27	0.43	0.03	0.02	0.16	0.18	0.18	0.26
1983	0.27	0.43	0.07	0.04	0.16	0.18	0.19	0.27
1984	0.27	0.43	0.09	0.04	0.16	0.18	0.20	0.27
1985	0.27	0.43	0.07	0.04	0.16	0.20	0.19	0.27
1986	0.27	0.43	0.06	0.03	0.16	0.18	0.19	0.27
1987	0.27	0.43	0.05	0.03	0.16	0.19	0.19	0.27
1988	0.27	0.43	0.05	0.02	0.16	0.19	0.19	0.27
1989	0.27	0.43	0.04	0.02	0.16	0.18	0.18	0.26
1990	0.27	0.43	0.04	0.02	0.16	0.22	0.19	0.27
1991	0.27	0.43	0.08	0.04	0.15	0.19	0.19	0.27
1992	0.26	0.43	0.12	0.08	0.14	0.16	0.20	0.26
1993	0.27	0.43	0.16	0.10	0.13	0.14	0.21	0.26
1994	0.26	0.43	0.15	0.14	0.13	0.16	0.21	0.26
1995	0.27	0.43	0.14	0.15	0.12	0.13	0.21	0.26
1996	0.27	0.43	0.16	0.12	0.11	0.13	0.20	0.26
1997	0.27	0.43	0.17	0.15	0.10	0.14	0.21	0.27
1998	0.27	0.43	0.14	0.15	0.10	0.13	0.21	0.26
1999	0.27	0.43	0.12	0.15	0.10	0.10	0.20	0.26

US	Scaled index of 9th to 1st decile	Scaled incidence of low-wage unemployment	Scaled standardized unemployment rate	Scaled long-term unemployment rate	Scaled injury rate	Scaled fatality rate	Scaled risk from poverty at end of life	Arithmetic average	Power average (alpha=3)
1980	0.73	0.73	0.28	0.04	0.98	0.34	0.40	0.50	0.26
1981	0.76	0.74	0.30	0.06	0.98	0.33	0.39	0.51	0.27
1982	0.78	0.75	0.39	0.09	0.98	0.32	0.39	0.53	0.28
1983	0.80	0.68	0.38	0.16	0.98	0.24	0.38	0.52	0.27
1984	0.82	0.72	0.30	0.12	0.98	0.28	0.37	0.51	0.28
1985	0.84	0.79	0.29	0.09	0.98	0.27	0.36	0.52	0.30
1986	0.88	0.78	0.28	0.08	0.98	0.26	0.35	0.51	0.31
1987	0.89	0.79	0.25	0.06	0.98	0.23	0.34	0.51	0.31
1988	0.91	0.80	0.22	0.05	0.98	0.22	0.32	0.50	0.32
1989	0.89	0.78	0.21	0.04	0.98	0.23	0.31	0.49	0.31
1990	0.85	0.77	0.22	0.04	0.98	0.18	0.30	0.48	0.29
1991	0.86	0.75	0.27	0.05	0.98	0.18	0.29	0.48	0.29
1992	0.87	0.77	0.30	0.10	0.98	0.18	0.28	0.50	0.30
1993	0.87	0.80	0.28	0.10	0.93	0.18	0.27	0.49	0.29
1994	0.94	0.84	0.24	0.09	0.91	0.18	0.26	0.49	0.31
1995	0.94	0.84	0.22	0.07	0.88	0.18	0.27	0.49	0.31
1996	0.94	0.84	0.22	0.06	0.81	0.18	0.29	0.48	0.29
1997	0.94	0.83	0.20	0.05	0.78	0.18	0.31	0.47	0.27
1998	0.94	0.82	0.18	0.05	0.73	0.18	0.31	0.46	0.26
1999	0.94	0.82	0.17	0.04	0.69	0.14	0.31	0.45	0.25

Increase in variables corresponds to a decrease in well-being consistent with HPI methodology. Data is taken from the same sources as in the Index of Labour Market Well-Being, and scaled in the same range as is used in the Index of Labour Market Well-Being, although scaling is not used in the HPI.

Typology of Methodological Choices Associated with Composite Index Construction

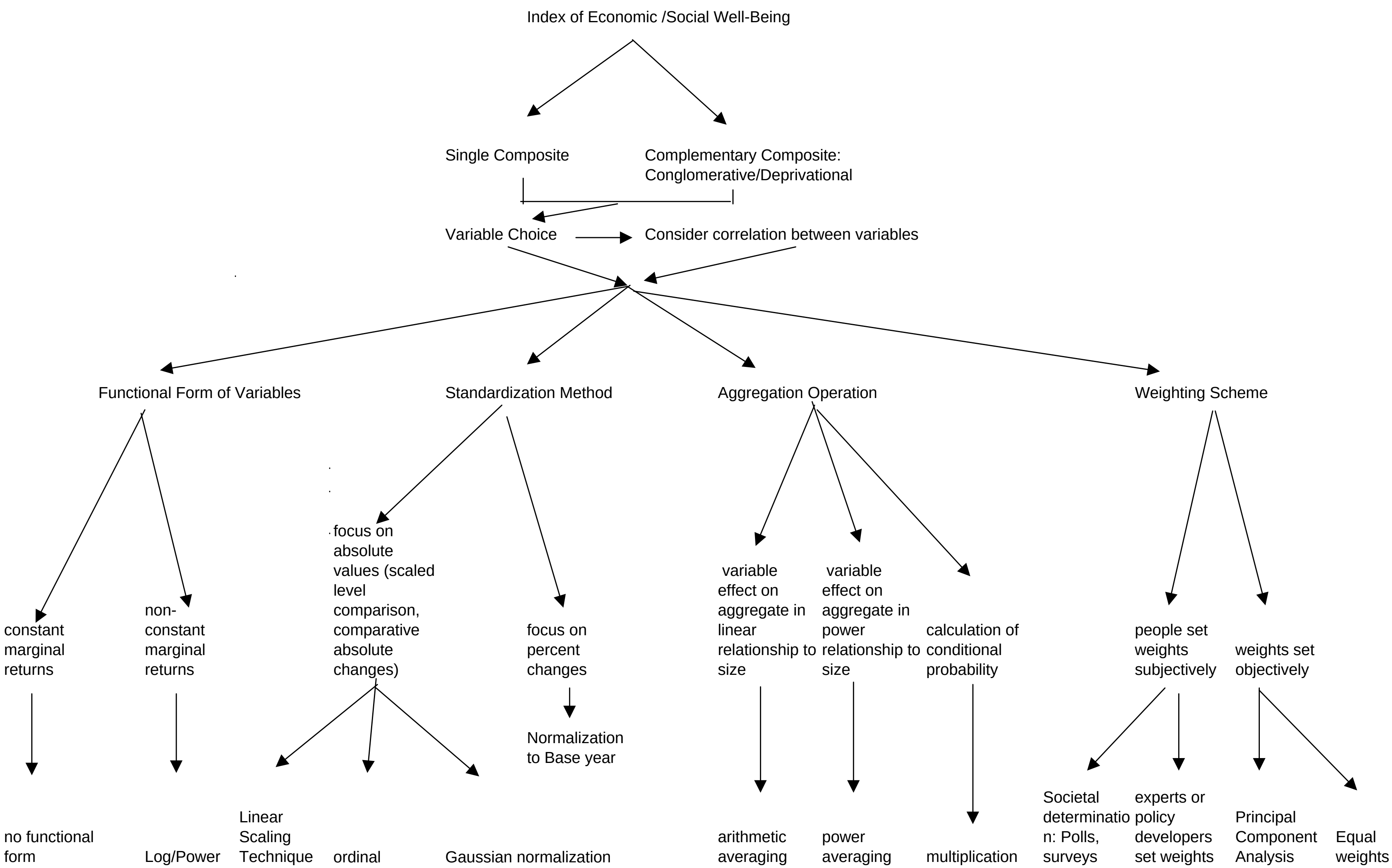
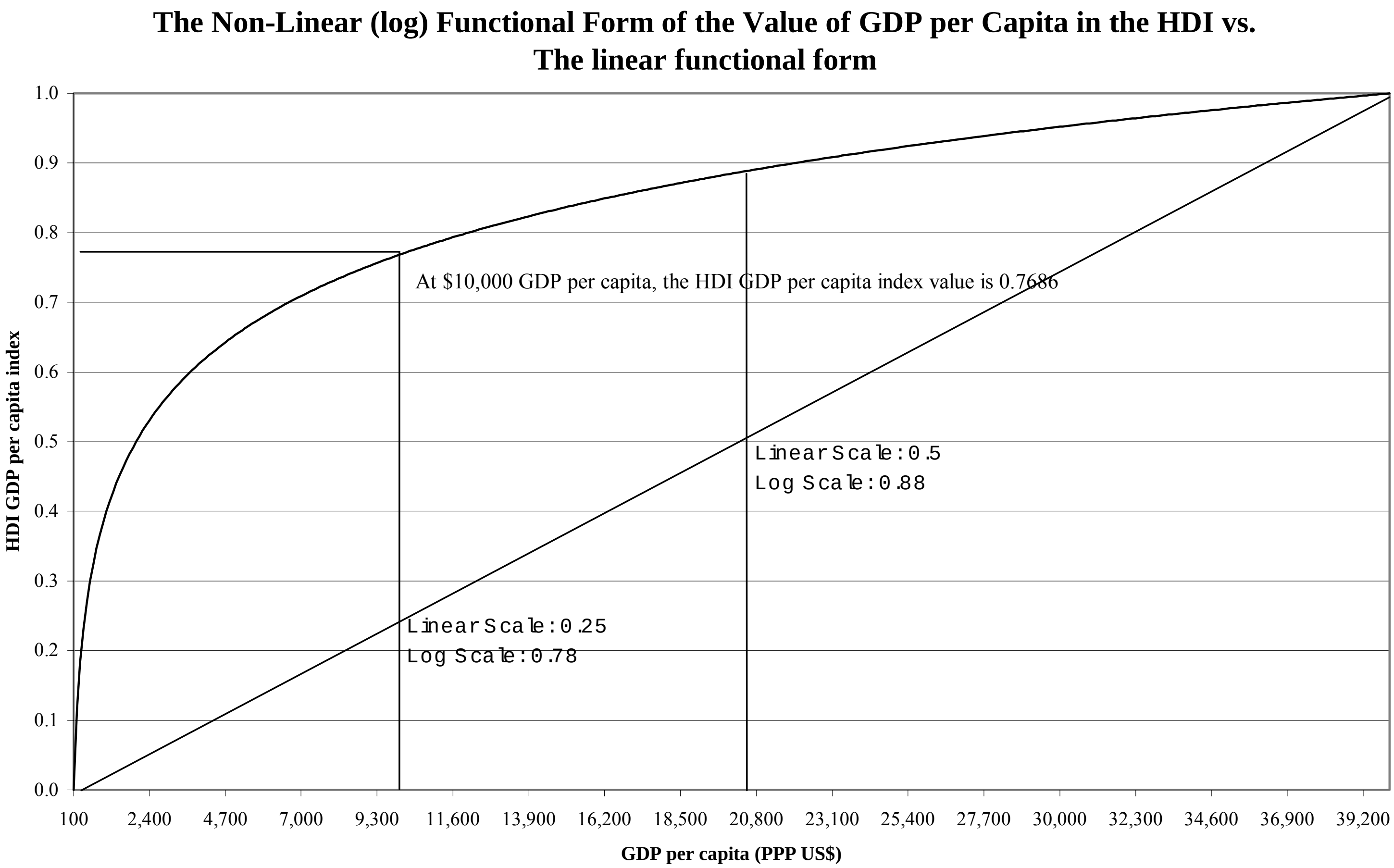


Chart 1: Level of linear and non linear function form for the GDP index of the HDI (for precise formatting, see the excel file gadrey-hdi in "my documents"



HDI GDP per capita index = (log(GDP per capita)-log(100))/(log(40,000)-log(100)).
Source: Human Development Report 2001, p. 240, United Nations.

Exhibit 2: Does the methodology address the issue or problem?

Functional Forms

	Linear	log/power
transparency	Yes	No
allow for varying marginal returns	No	Yes

Scaling of variables

	Linear Scaling (LST and Ordinal ratings)	Gaussian normalization	Normalization to base year
directionality issue	Yes	Yes	No
Large % range compared to base	Yes	Yes	No
Unequal implicit weights	Yes	No	No
Transparency	No	No	Yes
emphasis on percentage change or trend analysis	No	No	Yes

Aggregation Operation

	Arithmetical average	Power average	Multiplicative
Weighting in proportion to magnitude of component	No	Yes	No
Conditional Probability calculations	No	No	Yes
Transparency	Yes	No	No

Explicit Weighting

	PCA	Expert (including Developers)	Equal Weights[12]	Polls
Complete mathematical determination	Yes	No	No	No

[12] Equal weights means the decision to weight all variables equally, regardless of which variables are chosen and which values they take on. It is possible that any other technique will assign equal weights to variables, but this category means the explicit decision to assign equal weights.

Societal determination (polls)	No	No	No	Yes
Weights not subject to personal bias of index developers or experts	Yes	No	Yes	Yes
Weights set according to subjective considerations	No	Yes	No	Yes
Transparency	No	Yes	Yes	Yes

Exhibit 3: Summary of variables used in the Human Development Index (HDI), Human Poverty Index for developing countries (HPI-1) and Human Poverty Index for developed countries (HPI-2) (UNDP:2002):

Equal weighting is used in each index:

HDI

- Per capita GDP (first log is taken, then the quantity is scaled with LST)
- (one third)Literacy rate (%) (scaled with LST)+ (two thirds) Gross enrollment ratio: combined primary, secondary and tertiary enrollment (%)
- Life expectancy (scaled with LST)

HPI-1

- probability at birth of not surviving to age 40 (%times 100)
- adult illiteracy rate (%)
- unweighted average of population not using improved water sources (%) and underweight children under age 5 (%)

HPI-2

- probability at birth of not surviving to age 60 (%times 100)
- adults lacking functional literacy skill
- population below income poverty line (50% median disposable income)
- long-term unemployment rate (lasting 12 months or more)

References:

Adelman, Irema, Taft Morris, Cynthia, “A Factor Analysis of the Interrelationship Between Social and Political Variables and Per Capita Gross National Product. *The Quarterly Journal of Economics*, Volume 79, Issue 4 (Nov. 1965): 555-578.

Anand, Sudhir, and Sen, Amartya. “Human Development Index: Methodology and Measurement”. Occasional Paper 12, Human Development Report Office, 1994.

Anand, Sudhir and Sen, Amartya Gender Inequality in Human Development: Theories and Measurement. Occasional Paper 19, Human Development Report Office, 1995.

Anand, Sudhir and Sen, Amartya, “Concepts of Human Development and Poverty: A Multidimensional Perspective” *Human Development Papers 1997*, UNDP.

Anand, Sudhir and Sen, Amartya. The Income Component of the Human Development Index. *Journal of Human Development,* Vol. 1, no. 1, 2000.

Anderson, J.R., J.B. Hardaker, and J.L. Dillon, *Agricultural Decision Analysis.* Ames: Iowa State University Press, 1997.

Booysen, Frederik. An overview and evaluation of composite indices of development.
Social Indicators Research, 59(2), 2002(August): 115-151.

Brink, Satya and Zeesman,Allen. “Measuring Social Well-Being: An Index of Social Health for Canada”. June 1997 R-97-9E. http://www.hrdc-drhc.gc.ca/sp-ps/arb-dgra/publications/research/abr-97-9e.shtml

Chakravorty, Satya., “The variance as a subgroup of decomposition measures” *Social Indicators Research* 53(1), Jan 2001.

Deijain, Lai, “Temporal Analysis of Human Development Indicators: Principal Components Approach”. SIR 331-366 Sept. 2000.

Dunteman, George H. “Principal Components Analysis.” *Quantitative Applications in the Social Sciences*. No. 69 Sage University Paper. Sara Miller McCune, Sage Publications Inc. 1989

Gwartney, James and Lawson Robery. 2000 Annual Report Economic Freedom of the World. The Fraser Institute 2000.

Michael R. Hagerty, Robert A. Cummins, Abbott L. Ferriss, Kenneth Land, Alex C. Michalos, Mark Peterson, Andrew Sharpe, Joseph Sirgy, and Joachim Vogel (2001) “Quality of Life Indexes for National Policy: Review and Agenda for Research,” *Social Indicators Research* June, 2001, pp. 1-91.

Jolliffe, I.T.“Principal Components Analysis”. Springer-Verlag. New York, 1986.

Kelley, Allen C. 1991. "The Human Development Index: Handle with Care." Population and Development Review 17(2): 315-24.

Lawley, D.N. *Factor analysis as a statistical method*. Butterworth and co, 1971.

McGillivray, Mark and Howard White. "Measuring Development : The UNDP's Human Development Index", Paper presented for the ESRC Development Economics Study Group Annual Conference, March 27-28, University of Leicester, 1992.

Morton, Peter. “ Atlas of Canada QOL Model”. GeoAccess Divisions, Natural Resources Canada. 2003.

Noorbakhsh, Farhad. “The Human Development Index: Some Technical Issues and Alternative Indices. *J. Int. Dev*. 10, 589-605 (1998).

Noorbakhsh, Farhad. “A Modified Human Development Index”. *World Development*, Vol.36, No3. 517-528, 1998.

Lars Osberg and Andrew Sharpe. An Index of Economic Well-Being for Selected OECD Countries. *The Review of Income and Wealth. Income and Wealth*, series 48, no 3: 291-316. September 2002.

Shen, S.M., Lai, Y.L., “Optimally scaled Quality of Life Indicators.” *Social Indicators Research* 44(2) 225-254. June 1998.

Sinden, J.A. “Application of Quality of Life Indicators to Socioeconomic Problems: An Extension of Liu’s Method to Evaluate Policies for 26 Australian Towns”. *American Journal of Economics and Sociology*. 41 (4). October, 1982, 401-420.

Social Development in Pakistan. Annual Review 2001. Social Policy and Development Centre, Karachi. Oxford Press.

Srinivasan, T.N. “Human Development: A New Paradigm or Reinvention of the Wheel?”. *The American Economic Review*, Vol. 84, Is. 2, papers and Proceedings of the Hundred and Sixth Annual Meeting of the American Economics Association (May, 1994), 238-243.

Tatlidil, Huseyin. "A New Approach for Human Development: Human Development Scores", IDS, Sussex, Mimeo, 1992.

UNDP, *Human development Report 1993*, New York, Oxford University Press, 1993., and all years inclusive through 2002.

Vital Signs**.** Portsmouth, N.H. : Boynton/Cook, Heinemann, c1990-.

World Economic Forum, *1996 Global Competitiveness Report.*