ADRIAN OTOIU
The Academy Of Economic Studies, Romania

EMILIA TITAN
The Academy Of Economic Studies, Romania

AN ALTERNATIVE METHOD OF COMPONENT AGGREGATION FOR COMPUTING MULTIDIMENSIONAL WELL-BEING INDICATORS

Abstract:

There is considerable debate on the methods used to compute composite indicators of well-being. The fact that most of the weights of the principal sub-components of the composite indicators are equal, and that the determinants of well-being are, to a certain extent, correlated, makes the use of ranks of these sub-components in computing the country ranks of well-being indicators a valid approach. A comparison of the actual ranks with ranks computed as averages of the ranks of subcomponent indexes for three well-known indicators of well-being, Human Development Index, Legatum Prosperity Index, and Social Progress Index, shows that results are almost the same. This calls into question the use of weighted averages of actual values of sub-components, as very high values for a variable or sub-component increases a country's relative rank, despite much lower performance on other sub-components. Our proposed approach will help achieve more robust/reliable rankings of countries and tackle the issues posed by extreme values or non-normal distributions of the sub-components variables used.

Keywords:

well-being indexes, composite indices, rank-based statistical methods, Human Development Index, Legatum Prosperity Index, Social Progress Index, precision, recall

JEL Classification: C14, I30, C40

Introduction

The measurement of well-being and human progress has received an increasing attention in the recent decades. The need to measure human progress using multidimensional measures that take into account not only economic progress but also increased life expectancy, educational attainment, sustainability, etc., has occurred following the seminal work of development economists that have shown how differences in economic wealth do not automatically translate in a longer life expectancy, or higher levels of self-reported well-being/ life satisfaction. Triggered by the limitations of the GDP as an indicator of human progress, the quest for finding 'beyond GDP" measures has intensified in the recent years, with several new indexes being developed and made widely available by organizations led by prestigious members from industry, academia, media, think tanks, and non-profit sector. The Beyond GDP web site of the European Commission (2014) has links to over 15 indexes of well-being, which are constructed and disseminated by a wide variety of organizations, including statistical bodies, national agencies, charitable foundations, international organizations, and think tanks.

Among them, the Human Development Index (HDI), developed by UNDP is the most well-known. Since 1990 its rankings have been followed by media and governments and have also attracted significant academic interest with respect to its relevance. Among other indexes, the Legatum Prosperity Index (LPI), which appears since 2009, has eight sub-components that aggregate 89 variables (The Legatum Institute, 2012). The Social Progress Index's (SPI) first 2014 edition has three dimensions, each with four components, which all aggregate data from 54 indicators of social and environmental outcomes.

While the rankings and findings resulting from the actual computations of the indexes enjoy a large public debate and media attention, the methodological aspects behind computing these indexes also capture the attention of the academia with respect to their methodology. The HDI has enjoyed numerous critiques that referred to several aspects of it, among which its functional form, which resulted in a change in the way it is calculated. Other indexes aggregate across a large number of variables that capture specific aspects of well-being, using weights that are assigned following a process of normalization or standardization of variables.

While constructing these indexes is itself fraught with theoretical difficulties stemming from the fact that the indicator to be calculated represents a rather abstract and multidimensional concept, there is also the issue of the best methodology to be applied that will aggregate all dimensions of well-being in a meaningful and robust way. Often, we feel that there is a mismatch between the level of rigor that governs the treatment of the variables used in computing the indexes and the procedures for aggregating the indexes, and the actual results and interpretation of the rankings.

Therefore, we have explored whether there could be an alternative to the current calculation procedures of indexes and discovered that, in fact, a rank-based, nonparametric method of aggregating across the dimensions of well-being is a valid approach, which yields results comparable to the current ones. This approach can streamline the current methodology of calculating composite indexes and yield results that can accommodate non-normal distributions of the component variables.

In order to do this, we will first look at the composition of three indexes, HDI, LPI and SPI, and their methodology for computing and aggregating the component variables. Then, we will review the issues of variable aggregation and examine some of the major critiques of various methodological approaches. Next, we will present in detail our proposed methodology, provide the rationale behind it, and the potential advantages for using it. Lastly, we will test our proposed method by carrying out an analysis of the difference between the actual country ranks of the well-being indexes and the recalculated country ranks for the same indexes, and will draw conclusions and recommendations based on the findings.

Key Theoretical and Methodological Issues in Calculating Indexes of Well-Being

Devising an indicator that will enable us to measure human progress and well-being has a long history. One of the first salient attempts in this direction is the famous remark by Kuznets (1934, pp. 7), which observed that "The welfare of a nation can scarcely be inferred from a measurement of national income". Later on, the work of Amartya Sen (1999), the famous development economist, has shown that relatively affluent nations such as USA have not managed to better the lives of some of their citizens more than other, poorer, nations (e.g. Pakistan, India), did.

Among the most prominent initiatives that tackled the shortfalls of GDP as a measure of well-being, we note the Beyond GDP initiative of the European Commission (2014), the Stiglitz-Sen-Fitoussi (SSF) report (Stiglitz et al., 2009) and the OECD Better Life Index (OECD, 2014). All of them attempt to gauge the multidimensional aspect of well-being by identifying the indicators that address its many aspects, synthesize research findings in the field, and make recommendations as to which variables to choose and how to use them (SSF), or how to use the multiple dimensions of well-being in constructing the indexes (Better Life Index).

Apart from issues that involve data quality, reliability of the statistical estimates, and cross-country comparability, aggregation of variables into the indexes of well-being has posed challenges in terms of how to actually do that with respect to weighting their components, and how and whether to transform them so as to ensure that their cross-country variability is appropriately captured and factored in into the resulting composite indicators.

The basic variables used in the computation of the indicators must be transformed so as to make the different orders of magnitude and units of measure compatible with each other, in order to be aggregated using the formulas for composite indexes (OECD, 2008). This is a standard practice, called normalization of variables, which stands for several alternative procedures that can be performed on the raw data.

The first one, mentioned by OECD as being the simplest technique is ranking, through which the actual level of a variable is replaced by its relative position. It is criticized that it cannot reveal the absolute performance of a given country and that the information on the levels of the observed variable are not retained. Therefore, parametric techniques that take into account the actual values of the source variables and convert them using the parameters of the distribution (e.g. minimum value, maximum value, standard error, value corresponding to a reference country), are predominantly used in computing composite indicators.

Thus HDI components use the min-max technique (UNDP ,2012b), in which the relative position of a value of a variable is computed by comparing it to its minimum and maximum value using the formula 1, as shown below:

$$Ix_c = \frac{x_c - \min(x)}{\max(x) - \min(x)} \tag{1}$$

where Ix_c is the c element of the normalized (transformed) variable, x_c is the actual level of the initial variable, and max(x) and min(x) are the maximum and minimum observed values of the initial variable.

The methodology employed by the LPI in computing the variables that are built into the index is the z-scores (OECD, 2008), that convert the initial variables into scores with the mean of 0 and a standard deviation of one using the following formula:

$$Ix_c = \frac{x_c - \bar{x}}{\sigma(x)} \tag{2}$$

where Ix_c is the c element of the normalized (transformed) variable, x_c is the actual level of the initial variable, $\bar{x}$ is average of initial variable, and $\sigma(x)$ is the standard deviation, a measure of the variability of the x variable.

The basic variables used in constructing the SPI (Stern et al, 2014) were mostly scaled from 0 to 100 following a process of normalization performed by the providers of the data. Given this, we concluded that we can safely assume that most of them have undergone a parametric transformation of the raw data shown in the OECD methodological handbook for computing composite indicators (OECD, 2008).

The methodology for aggregating the data consists in either averaging across the components or aggregating the values of transformed variables using weights. For all indexes, the methodology comprises two main layers of aggregation, the first one where

variables are aggregated into main subcomponents and the second one where subcomponents are themselves aggregated to obtain the main composite index, HDI, LPI and SPI.

Aggregation methodologies are essentially consistent between the two layers. Thus, the HDI uses the geometric mean to aggregate the two education variables into the education index variable, then computes the index as the geometric mean of income, education and life expectancy indexes.

The LPI uses a more complex methodology, with income playing a role in the computation of each of the eight sub-components of the index: economy, entrepreneurship, governance, education, health, safety and security, personal freedom, and social capital. Thus, each sub-component is computed as the sum of the z-scores for income and well-being sub-indexes, which, in turn, are calculated as weighted coefficients of the normalized base variables (The Legatum Institute, 2012). Here the weights were obtained using the coefficients of regressions of the normalized income variables on the logs of GDP per capita in PPP (purchasing power parity adjustment) and, respectively, regressions of the normalized well-being variables on a dichotomous life satisfaction variable (The Legatum Institute, 2012). While weights were fairly diverse for the subcomponent variables, aggregation of subcomponents to obtain the LPI used equal weights of 1/8 for each sub-component.

The SPI determines the weights of aggregating the basic input variables into 12 components using factor analysis. Then, each component undergoes a standardization process which uses a version of the min-max technique to assign scores from 0 to 100 for each component. From this point up, aggregation of components into three dimensions, and aggregation of the three dimensions to obtain the SPI use equal weights (Stern et al., 2014).

The review of the methodologies used in obtaining the three well-being indicators thus exhibits the following main features:

-all input variables undergo a process of normalization, which uses a parametric method. In other words, the transformations of the raw variables with the purpose of aggregating them tends to retain the variability of the initial data (OECD 2008).

- weighting of several normalized variables uses factor analysis or regression analysis, in which the variability of each component variable in total data variability, or, respectively, the variability of a representative target variable, is quantified. These techniques are also parametric by nature, and their use is consistent with the use of variables normalized through using parametric methods.

- calculation of the components obtained by aggregating the variables uses equal weights for all variables, thus making the composite indexes LPI, HDI or SPI unweighted averages of all their component sub-indexes.

While from the methodological point of view these methodologies are sound and based on generally accepted practices, we believe that there are some methodological inconsistencies that advocate for the appropriateness of alternative ways for computing the composite indexes, that will simplify the current methodologies, and show increased flexibility in accommodating more subcomponents and variables, whose evolution may not affect our rankings as much as they should.

Are Parametric Methods Relevant And Appropriate For Computing Well-Being Indexes?

The case for parametric methods essentially rests with their ability to preserve the variability of the initial data, and convert the raw data from variables with different magnitudes into normalized variables, which are similar with respect to their unit of measurement and, in some cases, with their variability. For example, all HDI component variables and sub-indexes take values from 0 to 1, and all SPI components and dimensions take values from 0 to 100.

However, if we look at the methodology of computing the sub-indexes and composite indexes, we do realize that the components have equal weights, therefore the value of one sub-index does not appear to be more important than the value of another. This is to a large extent motivated by the fact that well-being is a complex measure of a rather abstract concept, which is very subjective, and influenced by multiple factors/variables. However, calculation of well-being must also take into account objective measures of well-being in order to measure 'what matters' (Michalos et al., 2011).

In the case of sub-indexes, correlation analysis shows the fact that sub-components are significantly related to each other for each and every index.

Thus, all SPI sub-components show a strong cross-correlation of above 70%. Similarly, HDI components show a strong correlation between life expectancy at birth and education, and a medium-to strong correlation between of these two-sub-indexes and the GNI per capita.

Table 1. Correlations between SPI components

	Basic Human Needs	Foundations of Wellbeing	Opportunity
Basic Human Needs	1		
Foundations of Wellbeing	82%	1	
Opportunity	70%	85%	1

Source: authors' calculation

Table 2. Correlations between HDI components

	Life Expectancy at Birth	Education Index	Gross National Income (GNI) per capita
Life Expectancy at Birth	1		
Education Index	76%	1	
Gross National Income (GNI) per capita	60%	60%	1

Source: authors' calculation

Table 3. Correlation between LPI components

	Economy	Entrepreneurship	Governance	Education	Health	Safety & security	Pers. freedom	Social capital
Economy	100%							
Entrepreneurship	77%	100%						
Governance	71%	86%	100%					
Education	71%	91%	75%	100%				
Health	75%	94%	81%	93%	100%			
Safety & security	61%	84%	79%	81%	84%	100%		
Pers. freedom	55%	56%	66%	47%	51%	64%	100%	
Social capital	65%	67%	56%	66%	64%	58%	53%	100%

Source: authors' calculation

In the case of LPI, we observe more complex relationships given by the heterogeneous nature of its sub-components. Correlations between the Economy sub-component at the other sub-components are strong and medium-to strong, over 60% for all but one of them. Similar results are obtained for other components, except for Personal freedom and Social capital. For them, cross-correlations are in the medium-to-strong range; however, with one exception, none of the correlations fall under 50%.

Thus, given a significant degree of uncertainty of how well-being is to be measured, and given the fact that there is a significant degree of correlation between all its sub-component measures, we do not consider the fact that a higher absolute value of a given subcomponent should decisively affect the ranking of a country given by a composite well-being indicator. Consequently, we question the actual gain and relevance to be obtained through using variables or components which are based on parametric normalization methods.

Furthermore, the high variability of the composite measures of some sub-indexes for some countries does not appear to bear much relevance with respect to their overall level

of well-being. More precisely, well-being is expressed as country ranks, and almost all of the discussions concerning the relative evolution of a given country are held in terms of its relative position among all nations for which well-being indexes are computed. Thus, most of the analyses comment the fact that a country has improved in terms of ranks, but the fact that their relative subcomponent score has changed from value x to value z is a rather technical and abstract argument, which only makes sense, and is to be ultimately explained, by the evolution of the initial variables.

Let's take the example of the top 2014 SPI ranked country, New Zealand, which comes in first despite being ranked 18 on the basic human needs dimension, and the fact that all of the four runner-ups have scores for the two other dimensions above it. A similar situation is observed for the US, where a third rank on the 2012 HDI seems not to be affected by a rather mediocre performance on life expectancy, which compares the ones observed for countries having overall ranks of 15 and below.

The non-parametric methods, that are very seldom used (OECD, 2008) or appreciated in the context of computing well-being indicators, are widely used in statistical analyses in many fields. Tests of relevance and correlations use rank-based methods to compare variables, and are widely used in a variety of statistical applications. Their particular strength is given not only by their simplicity of calculation, but also by their ability to deal with data exhibiting non-normal distributions, which mitigates the impact of some extremely high or low values, or a lack of existing data for particular intervals of the distribution of a variable.

Comparisons between parametric and non-parametric methods have shown that non-parametric methods are more appropriate when the underlying assumptions governing the parametric tests are not met, and that they are viable alternatives when data does not enable a 'metric' or data-specific interpretation that is highly dependent on the very levels of the analyzed variables (Harwell, 1988). The Wilcoxon-Mann-Whitney non-parametric test appear to perform better than the t-test in the case of non-normal distributions with respect to the type I error and to the heterogeneity of variances (Zimmerman, 1998). Recent comparisons between parametric and non-parametric tests showed that they can perform better in models for some sensitive markets (Hernandez and Torero, 2013).

Therefore, given the properties and their performance, and given the fact that there is not a strong case for preserving the initial variability of the data in order to explain variations of the composite indexes well-being, we consider that a non-parametric approach may be more suitable to rank the well-being of countries starting from the ranking of sub-indexes. We consider that such a method will take care of the strong variation in the normalized variables that make up the sub-indexes, and produce results that are more compatible with the multidimensional nature of the composite indexes, which mitigate the influence of the potentially large variation in the variables or sub-components used in computing the relative country ranks.

Recalculating the Well-Being Indexes Using Rank Aggregation of Sub-Indexes

We have tested the validity of using ranks for computing indexes of well-being by recalculating the country ranks of the well-being indicators based on the ranks of the component sub-indexes. The relative ranks of sub-components were summed up, and the resulting rank for the composite indicators was computed starting with the lowest sum of the variables as shown in formula 3

$$RC_i = rank_i(\textstyle\sum_j^k rank_{ij}) \tag{3}$$

where RC stands for the rank of country, and i, and k denotes the number of sub-indices to be aggregated.

Results from recalculated ranks, summarized in table 4, were similar to the actual ones obtained using parametric methods. In over 2/3rds of the cases, absolute rank differences did not differ by more than five ranks, with only a handful of cases showing a rank difference above 10. Differences were most prominent for HDI, which uses the least number of variables for computing the sub-indexes of well-being. The SPI and LPI showed similar results with less than 16 % and, respectively, 9% of the countries showing an absolute rank difference bigger than five. A brief comparison of the countries that show the greatest rank differences reveal the fact that, as shown in Appendix 1, the differences are mainly random.

Table 4. The distribution of rank differences between original index ranks and recalculated ranks

Rank differences (actual-recalculated)	HDI	SPI	LPI
< -14	0.5%	0.0%	0.0%
-14 to -10	2.7%	0.0%	0.7%
-9 to -5	12.8%	6.8%	4.2%
-4 to 0	43.1%	54.5%	64.8%
1 to 5	28.2%	31.8%	26.8%
6 to10	9.0%	6.1%	3.5%
11 to 15	3.7%	0.8%	0.0%

Source: authors'calculations

In order to further compare the classifications obtained following the recalculation of ranks, we have assessed the country classifications from the three indexes of well-being against the classifications of countries based on our recalculations, following the same

classification rules laid out in their dissemination reports (UNDP, 2013, Porter and Stern, 2013, The Legatum Institute, 2012).

Based on the obtained rankings, all of the three indexes have provided a broad classification of countries into groups, as shown in table 5.

The HDI and LPI have split the rated countries into four groups that are fairly homogeneous and delimited in a rather consistent way. Thus, the highest and lowest LPI ranks group 30 countries each, with the middle tiers grouping 41 countries each. HDI has a more homogeneous grouping, with equal sized groupings split at the 47th, 94th and 141th rank. By contrast the SPI splits are uneven, with highest and lowest ranks grouping a rather small number of countries (less than 20), and middle to low tiers grouping most of the countries.

Table 5. Country classification by the composite indexes of well-being and its mapping

Mapping	HDI	SPI	LPI
A	Very High Human Development	Top 10	High Ranking Countries
B	High Human Development	Next tier	Upper Middle Ranking Countries
C	Medium Human Development	Third tier	Lower Middle Ranking Countries
D	Low Human Development	Fourth tier	Low Ranking Countries
E	-	Fifth tier	-
F	-	Bottom tier	-

Source: The Legatum Institute (2012), UNDP (2013), Stern, S., Wares, A. Orzell, S.(2014)

Comparison between actual ranks and recalculated ranks was carried out using the precision and recall, two measures of classification accuracy stemming from information theory. In order to assess the extent to which the country classifications using the recalculated ranks of countries match the existing classifications, we have compared the two classifications for each country and grouped the results into the following categories, based on the method described by Hand et al (2001):

- True positives: when recalculated classification match the original classification

- False positives: when recalculated classification correspond to different original classification

- False negatives: when original classification are not matched by the recalculated classification

The formulas for computing precision and recall are:

Precision=true positives/(true positives+false positives)   (4)

Recall=true positives/(true positives+false negatives)   (5)

The intuitive, non-technical explanation of these concepts is as follows:

- Precision tells to what extent countries that belong to a particular class after reclassification match the original members of that class

- Recall tells to what extent the original members of a class have grouped in the same class after reclassification

The interplay between the two measures is very important as it corresponds to two important aspects of the quality of a reclassification. In our case, it shows how the reclassification of countries based on the ranks of the sub-indexes matches the existing classification. The first aspect, captured by precision, shows how many of the countries predicted to fall in a given rank have belonged to that rank in the original classification. While valuable, precision is not sufficient by itself as it cannot tell how many of the countries grouped in the original classification fall in the same classes after reclassification is carried out. This aspect is shown by recall.

Using the groupings obtained from the recalculated ranks of the countries, we have gotten the following results:

- For HDI (Table 6) , precision and recall yield an excellent score of 95.7%, with 95.7% for group A, 91.5% for group B, 95.7% for group C, and 100% for group D. One aspect to be noted is the fact that, due to the symmetry of the misclassifications, results for precision and recall are identical.

- For LPI (Table 7), precision and recall yield an excellent score of 95.8%, with 96.7% for group A and D, and 95.1% for group B and C. Again, due to the symmetry of the misclassifications, results for precision and recall are identical.

Table 6. Precision and recall for HDI

Predicted groups

Actual groups	A	B	C	D	Total
A	45	2			47
B	2	43	2		47
C		2	45		47
D				46	46
Total	47	47	47	46	187

Table 7. Precision and recall for LPI

Predicted groups

Actual groups	A	B	C	D	Total
A	29	1			30
B	1	39	1		41
C		1	39	1	41
D			1	29	30
Total	30	41	41	30	142

For SPI (Table 8), precision and recall yield an excellent score of 93.9%, with 90% for group A, 84.6% for group B, 87.5% for group C, 96.2% for group D, 97% for group E and 100% for group F. A brief inspection reveals the fact that the scores lower than 90% are in fact due to the small number of elements within a group, thus making relatively small misclassifications of one to two elements appear relatively larger for groups consisting of under 20 countries (B, C) compared to groups of 40 countries and over (D and E).

Table 8. Precision and recall for SPI

Predicted groups

Actual groups	A	B	C	D	E	F	Total
A	9	1					10
B	1	11	1				13
C		1	14	1			16
D			1	50	1		52
E				1	32		33
F						8	8
Total	10	13	16	52	33	8	132

Conclusion

Using ranks in computing indexes of well-being proves to be a valid aggregation method, yielding reliable results.

The recalculation of country ranks for three of the most well-known indices of well-being, the HDI, LPI and SPI, using the ranks of their sub-indexes, has yielded results that were highly comparable with the original calculations. High precision and recall measures, combined with the fact that the vast majority of countries ranks do not differ by more than five ranks, show that the use of ranks for computing composite indicators is a valid alternative approach yielding results close to the ones obtained with parametric methods.

Combined with the fact that sub-indexes are aggregated using equal weights, and that there is no established theory or opinion to show that the strong evolution of initial variable, or sub-component, in absolute terms, can and should compensate the opposite evolution of other variables in the context of defining multi-dimensional well-being, we can conclude that our method is not only a statistically valid technique, but one that is also sound from the methodological point of view.

The fact that ranks do not preserve the absolute variation of the component variables is compensated by the fact that ranks are more effective in dealing with outliers and with asymmetric distributions. In this respect, it is our belief that using ranks for the sub-component indexes, is better, as it can factor in the relative overall performance of a country by taking into account the ranks and not the absolute values of the sub-indexes. This latter practice, used by all three indexes, leads to relative ranks that are inflated by a strong performance in some sub-components and/or mitigate the poor performance on other sub-components.

Given the fact that country ranks produced by these indicators receive significant media attention and submit the governments of the ranked countries to scrutiny, current methods of computing well-being indicators may overemphasize the achievements of the governments and conceal relative weaknesses of indicators. For example, New Zealand's top position in the SPI ranking, conceals a rank of 18 in the basic human needs component. While the reclassified SPI ranking keeps the country in the top 10, we cannot fail to notice the fact that no other country in the initial top 10 has a rank as low as 18 in any of the three sub-components. Such a partial underperformance is more likely to get noticed when the country is close to the bottom of the top 10 countries compared to the case when it got the top position among 132 countries. A similar case can be made for the USA, whose its third position in the HDI ranks conceals a 35th rank on life expectancy, significantly lower than life expectancy ranks of any other of the top 10 ranked countries.

Appendix 1

	HDI			SPI			LPI	
actual rank	Country	rank difference	actual rank	Country	rank difference	actual rank	Country	rank difference
59	Cuba	11	1	New Zealand	-7	84	Georgia	8
59	Panama	10	67	Venezuela	7	92	Laos	8
50	Belarus	-13	89	Mongolia	11	83	Albania	-6
55	Russian Federation	-10	88	Indonesia	7	104	Senegal	6
52	Palau	-15	96	Ghana	7	109	Niger	7
87	Armenia	14	110	Congo, Rep.	10	101	Iran	-10
92	Sri Lanka	15	102	India	-9	131	Côte d'Ivoire	8
72	Saint Kitts and Nevis	-10						
72	Lebanon	-12						
104	Maldives	11						
83	Saint Vincent and the Grenadines	-11						
116	Syrian Arab Republic	11						

References

European Commission (2014). Beyond GDP. URL address: http://www.beyond-gdp.eu/, Accessed June 21, 2014

Hand, D., Mannila, H., Smyth P. (2001). Principles of Data Mining. The MIT Press, chapter 14, p770

Harwell, M. (1988). Choosing between Parametric and Nonparametric Tests, Journal of Counseling and Development: JCD67.1 (Sep 1988): 35

Hernandez, M. A., & Torero, M. (2013). Parametric versus nonparametric methods in risk scoring: an application to microcredit. Empirical Economics, 46(3), 1057–1079. doi:10.1007/s00181-013-0703-8

Kuznets, S. (1934). "National Income, 1929–1932". 73rd US Congress, 2d session, Senate document no. 124, page 7. http://library.bea.gov/u?/SOD,888

Michalos, A.C., Smale, B., Labonté, R., Muharjarine, N. (2011). The Canadian Index of Wellbeing. Technical Report 1.0. Vasa. Retrieved from http://medcontent.metapress.com/index/A65RM03P4874243N.pdf on February 6, 2013

OECD (2014), OECD Better Life Index http://www.oecdbetterlifeindex.org/, accessed November 14, 2012.

OECD (2008). Handbook on Constructing Composite Indicators: Methodology And User Guide

Porter M. E., Stern S. (2014). Social Progress Index 2014

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

Stern, S., Wares, A. Orzell, S.(2014). 2014 Social Progress Index, Methodological Report

Stiglitz, J., Sen, A., & Fitoussi, J.-P. (2009). Report by the commission on the measurement of economic performance and social progress. Paris, France: Commission on the Measurement of Economic Performance and Social Progress.

The Legatum Institute (2012). The 2012 Legatum Prosperity Index

UNDP (2012a). Human Development report 1990-2011. available at 862 http://hdr.undp.org/en/, Accessed Nov 30 2013. 863

UNDP   (2012b).   Human   Development   Index   Technical   Notes.   864 http://hdr.undp.org/en/media/HDR_2011_EN_TechNotes.pdf

UNDP (2013). Human Development Report 2013, The Rise of the South: Human Progress in a Diverse World

Zimmerman, D. W. (1998). Invalidation of Parametric and Nonparametric Statistical Tests by Concurrent Violation of Two Assumptions. The Journal of Experimental Education, 67(1), 55–68.