
DEPARTMENT OF ECONOMICS

Bachelor Essay in Econometrics, 2010

# A new index of welfare

## – An econometric approach on a new index of welfare generated with principal component analysis

Supervisor:

Fredrik NG Andersson

Author:

Johan Edfeldt

# Abstract


**Title:** A new index of welfare – An econometric approach on a new index of welfare generated with principal component analysis

**Topic/course:** NEKK01

**Author:** Johan Edfeldt

**Supervisor:** Fredrik NG Andersson

**Keywords:** Welfare, Well-being, Principal Component Analysis (PCA), Cluster analysis, 1960-2008

**Purpose:** The purpose of this essay is to generate a wider measurement of welfare by carefully chosen variables in mainly three different areas: economic, socio-economic and environment in order to generate a welfare index which describes the world yesterday, today and tomorrow.

**Method:** With the Principal Component Analysis 14 variables from 115 countries are studied and verified and analyzed with regression models, cluster analysis and time series analysis.

**Conclusion:** In the present of 2008 the Generated Index explains the HDI and DI in a fairly good way thus it has the possibility to replace the two of them. However in a longer period of time the index is not to be equal to HDI. As for the DI there is only data for a couple of years and conclusion of this cannot be drawn during a longer period of time. The Generated Index could instead perhaps be used as a crisis predictor of some sort. The correlation with the different crisis during the time period 1960-2008 is imminent and with some extended study of the model a better crisis predictor can possibly be created.

## Table of Contents

# 1 Introduction

Is it possible for one index to describe the welfare of a nation which several other indexes aims to do? Could this index of welfare tell us more than just about the current situation in world and describe when there were crises and when there will be crises?

Today there exist several different measures of the welfare and well-being of a country such as: Human Development Index (HDI), Democratic Index (DI), Happy Planet Index (HPI), Index of Sustainable Economic Welfare (ISEW), Gross National Happiness (GNH), Genuine Progress Indicator (GPI), Physical Quality of Life Index (PQLI) et cetera. All of these indexes includes different variables in order to define what they argue is welfare and/or well-being.

The definition of welfare is according to Merriam Webster (Merriam-Webster 2010) "the state of doing well especially in respect to good fortune, happiness, well-being, or prosperity" and according to Dictionary.com (Dictionary.com 2010) "the good fortune, health, happiness, prosperity, etc., of a person, group, or organization; well-being: to look after a child's welfare; the physical or moral welfare of society." The definition of welfare is both broad and connected to well-being which is a more subjective way of the welfare concept.

Several of the indexes have variables chosen more or less arbitrary and ad hoc or an incomplete amount of variables. Other indexes have an incomplete diverse set of variables which results in the fact that the indexes differ greatly from each other. Instead of having several different indexes in order to find out what country has the best welfare one more advanced index would be superior.

The purpose of this essay is to generate a wider measurement of welfare by carefully chosen variables in mainly three different areas: economic, socio-economic and environment in order to generate a welfare index which describes the world yesterday, today and tomorrow.

In order to create a new index several things are needed. First of all the definitions of the term welfare and the term well-being needs to be discussed from different perspectives. This is conducted from a microeconomic framework digging deeper into consumer and producer surplus and of course the welfare function. From the welfare function the step is not too far to the social welfare function which is later discussed. That's the welfare part but what about the well-being? Are there similarities between welfare and well-being or are they actually two different

concepts and can well-being be measured at all? This part is trickier seeing that it is a subjective matter and it cannot be defined as clearly as the welfare. If the well-being is a measurement of happiness the argument loops to instead be a matter of defining happiness, a somewhat impossible quest. Seeing that welfare and sometimes well-being is argued about from a GDP (gross domestic product) point of view which is closely connected to consumption the concept of consumption is in need of being discussed.

As mention earlier there is a flood of different indexes claiming that they and they alone can measure welfare, well-being or both and in order to accept or reject them a study of their advantages and disadvantages is needed and undertaken.

Probably the most difficult task is to choose the variables of interest. They need to be well-spread and plenty, however a large number of variables soon become too complex to keep in mind while conducting the study. Mainly three categories of variables have been chosen: economic, socio-economic and environment aimed at capturing several different aspect of welfare. One interest of the study is to replace the HDI and DI with a new index and to do this the concept of democracy needs to be clarified and this essay will argue from a trade perspective. If this model shows to be successful a future study with more and different variables can be conducted hopefully to better describe the welfare in the world.

When the theoretic foundation has been laid the method of choice is discussed and the essay uses Principal Component Analysis (PCA) which aims to find and extract the most important information from the data and compress the size while at the same time keeping the important information and simplify the description of data. With the help from PCA an index can be created and reformed to a number between 0-100 where 100 is the country with the best welfare (the unreformed index is also used later, in the time series).

The results of the PCA can be verified in several ways and this essay uses something called biplot which is a graphical interpretation of the PCA to see if the variables and countries correlate in certain ways. This could be interesting in order to see if some countries cluster together with each other close to a specific variable and interesting conclusions can fairly easy be drawn.

Another method of verifying the PCAs results is dendrogram which supplements the biplot and has a possibility to cluster groups of countries together to see if, for example, all European countries groups together or not.

All essays needs a hypothesis and for this study it is that industrially developed and democratic countries will have a high Generated Index of Welfare (GI).

After the method is concluded the empirical analysis takes place and the data for this study is gathered from mainly three different sources: World Development Indicators Online (WDI) which is the World Bank's premier annual compilation of data about development and the global economy, the United Nations Development Programme (UNDP) which is the UN's global development network and The Economist Intelligence Unit's Index of Democracy 2008.

After the actual study the results can be presented and analyzed. And it was possible to generate an index which fulfills the hypothesis in a good way; however the stability of the model is in need of testing. This is done by creating the index with fewer principal components than all of the 14 to see if the same index is generated once more, which is the case in this study. Linear regressions are performed to find connections between the HDI and the DI as well as finding out the Generated Index actually brings something new to the field of welfare by comparing it with GNI (gross national income).

To be a useful index it needs to be able to describe the welfare over time which is why eleven countries, from the studied 115, are chosen from the top, middle and bottom of the Generated Index list to study how the vary over time. Together with the time series a crisis correlation is undertaken to see if and how the indexes is influenced by both national and international crisis. Sadly the statistics for several countries are flawed and not available from 1960 and forward which has to be taken into account when studying the figures with time series.

The conclusions are usually the last part of an essay and there are several which can be drawn. At the present of 2008 the Generated Index has the possibility to replace both the HDI and DI in a fairly good but over time this is probably not possible. The index can explain that the amount of crises and the instability of the welfare are more distinct in the period from the middle of the 1980s and forward. But the model itself has been proven well for this type of study and if the variables are changed and more complete statistics is found it should be possible to create a superior welfare index over time.

# 2 Theory

## 2.1 A microeconomic approach on welfare and well-being

An economist's use of welfare refers to the well-being of various groups such as consumers and producers. The analysis of the impact of a change on various groups' well-being is what constitutes welfare economics. Both policymakers and economists are interested in knowing how much consumers benefit from or are harmed by shocks that affect the equilibrium price and quantity. In order to find out the answer to this question there is a need for a welfare measurement tool. The consumer welfare is measure in terms of dollars. In other words how much dollars are gained or loss when for example a new tax on pollution is implemented.

Consumer welfare is the difference between what is gained and paid for a certain good. The consumer welfare is the pleasure a consumer get from the good above and beyond its price. For example if you were to buy an umbrella you are willing to pay more for it on a rainy day than on a sunny day not because it's a different good but because you get more pleasure out of it on a rainy day when it is used.

The tool for measuring welfare is the demand curve in which the consumer's marginal willingness to pay is reflected (i.e. the maximum amount a consumer will spend for an extra unit). In other words the consumer's marginal willingness to pay is the marginal value the consumer places on the last unit of output. The difference between the cost of a good and what the consumer is willing to pay for the good is the consumer surplus (CS) which is equal to the area under a demand curve.

The producer's welfare is measured by the producer surplus (PS) which is the difference between what the seller is willing to produce the good for and how much the good is sold for. Instead of using the demand curve the supply curve is used to measure the producer surplus.

The society's welfare in general is the sum of consumer surplus and producer surplus, Equation 1:

$$W = CS + PS \qquad \textbf{Equation 1}$$

This equation weights the well-being of consumers and producers equally thus implying that the well-being of producers and consumers are equally important.

The welfare lowers if the producing is less or more than the competitive output and competition maximizes the welfare because price equals marginal cost at the competitive equilibrium.

The tools for measuring welfare are needed to predict the impact of government policies and other events with the power of altering a competitive equilibrium. (Perloff 2009)

### 2.1.1 Arrow's Impossibility Theorem

Even if it is interesting to compare each and every ones specific preferences it is most difficult and that is why some sort of aggregation is needed. However Arrow's Impossibility Theorem shows that there is no ideal way to aggregate individual preferences into social preferences. The theorem requires a list of some things that a social decision mechanism should do (Varian 2006):

1. Given any set of complete, reflexive, and transitive individual preferences, the social decision mechanism should result in social preferences that satisfy the same properties.
2. If everybody prefers alternative x to alternative y, then the social preferences should rank x ahead of y.
3. The preferences between x and y should depend only on how people rank x versus y, and not on how they rank other alternatives.

The Arrow's impossibility Theorem is formulated as follows:

"Arrow's Impossibility Theorem: If a social decision mechanism satisfies properties 1, 2, and 3, then it must be a dictatorship: all social rankings are the rankings of one individual."(Varian 2006)

### 2.1.2 The welfare function

A person $i$ prefers $x$ to $y$ if and only if $u_i(x) > u_i(y)$ and one way of getting social preferences from individuals' preferences is to add up the individual preferences and use them as a kind of social utility. By doing this the allocation $x$ is socially preferred to allocation $y$ if (Equation 2)

$$\sum_{i=1}^{n} u_i(x) > \sum_{i=1}^{n} u_i(y) \qquad \textbf{Equation 2}$$

where $n$ is the number of individuals in the society. This is of course a totally arbitrary description since the choice of utility representation is totally arbitrary. A possible restriction on alloca-

tions above is that the "aggregating function" is to be increasing in each individual's utility. This will assure that everybody prefers $x$ to $y$, then the social preferences will prefer $x$ to $y$. The name for this kind of aggregation function is a social welfare function which in turn is just some function of the individual utility functions: $W\big(u_1(x), \dots, u_n(x)\big)$. However this gives a way to rank different allocations which depends only on the individual preferences and it is an increasing function of each individual's utility.

A form of this social welfare function is the classical utilitarian (or Benthamite welfare) function; the sum of individual utility functions, Equation 3:

$$W(u_1, \dots, u_n) = \sum_{i=1}^{n} u_i \qquad \textbf{Equation 3}$$

where $n$ is the number of individuals in the society.

The weighted sum-of-utilities welfare function is a slight generalization of this form, Equation 4:

$$W(u_1, \dots, u_n) = \sum_{i=1}^{n} a_i u_i \qquad \textbf{Equation 4}$$

where the weights, $a_1, \dots, a_n$ are supposed to be numbers indicating how important each agent's utility is to the overall social welfare (each $a_i$ is taken to be positive). In this essay the different weights will be generated from the principal component analysis in a non-arbitrary way in order to make them more trustworthy.

There are more welfare functions and another is the minimax or Rawlsian social welfare function which says that the social welfare of an allocation depends only on the welfare of the worst of agent - the person with the minimal utility, Equation 5:

$$W(u_1, \dots, u_n) = \min\{u_1, \dots, u_n\} \qquad \textbf{Equation 5}$$

The restriction on the structure of the welfare function at this point is that it be increasing in each consumer's utility. The problem of maximizing the welfare function is formulated and discussed at paragraph A1 Maximizing the welfare function at page 58 in the Appendix. (Varian 2006)

#### *2.1.2.1 Individualistic Social Welfare Functions*

A special form of welfare function known as an individualistic welfare function (Bergson-Samuelson welfare function) is directly a function of the individuals' utility levels and indirectly a function of the individual agents' consumption bundles. If $x_i$ denotes individual $i's$ consumption bundle and $u_i(x_i)$ is individual $i's$ utility level using some fixed representation of utility a social welfare function will be, Equation 6:

$$W = W\big(u_1(x_1), \dots, u_n(x_n)\big) \qquad \textbf{Equation 6}$$

There will be no consumption externalities if each agent's utility depends only on his or her own consumption.

## 2.2 Is well-being equal to welfare and/or is it possible to measure well-being at all?

It is often difficult to keep the terms well-being and welfare apart and for the most part they are both closely connected to each other. While well-being is a subject measurement, welfare aims to be more objective but this is not always the case. The different sources below discuss the concept of well-being.

Well-being is a subjective measure of the satisfaction or happiness of individuals and is not as clearly defined in economic as the term welfare. They are used synonymously in some studies and interpreted broader than welfare in other studies. (Kulig, Kolfoort et al. 2010) Well-being also lacks a fixed definition within psychology were the subjective well-being is thought to be comprised of four dimensions: life satisfaction, happiness, affect, and quality of life. (Steel, Schmidt et al. 2008)

Well-being is regarded as a multi-dimensional quantity, depending not only on what individuals have achieved or attained in the way of goods and services, but also on various attributes they enjoy as citizens (e.g. freedom, tolerance). (Sen 1987)

It all depends on what is included in the utility function but well-being is a subjective measure broader than the utility derived from the consumption of goods and services. (Kulig, Kolfoort et al. 2010)

Probably it would be of great interest to measure them both (welfare and well-being) at the same time but there are several problems with this. Several questions arises around the concept "Is it possible to measure well-being?", "Is welfare more important than well-being?", "Are welfare and well-being of equal importance?", "Is it possible to weight welfare different compared to well-being?" and "If the two concepts is not of equal importance then how do we weight them against each other?" et cetera. There is great risk with arbitrary in a case like this and should not be looked upon with casual eyes. In some way trying to measure well-being is like trying to measure happiness and this is most certainly impossible. To argue that it is possible to measure happiness the concept needs to be defined and to define happiness is like trying to define something that is different for each and every one of us. Of course a form of outline to happiness can be conducted but an outline is not nearly good enough to define it properly and to use it in measurement. This is questions with no clear answer and the problems surrounding this fact need to be considered continuously throughout the discussion of welfare and well-being.

## 2.3 Can consumption be equal to well-being?

Something as easy as consumption turns out to be fairly difficult to define. Consumption as a concept is different depending on which point of view one uses in the actual consumption. The person who buys a commodity, a good, has his or her own decisions to buy such as the concept of advertising, lifestyle and notion on consumer choice. Especially the notion on consumer choice is in itself in some way sacrosanct, encouraged and reproduced by the mass media which in turn spawns more detailed questions of consumption in everyday life. The other point of view is the commodity, the good, which is made up by material, social and cultural formations and it connect the world in different ways and links to specific resources, people, places and interests are a fact. (Paterson 2006)

To be able to consume one needs something to consume with, for example money. A way to measure the consumption possibilities is to look at the total value produced within a country, in other words the gross domestic product (GDP). The consumption is normally the largest part of the GDP and together with the income received from other countries (with interest and dividends) and payments made to other countries the gross national income (GNI) is formed. GNI is used in this essay as a form of consumption possibility measurement.

To argue that a high consumption possibility is equal to well-being is to argue that money can buy happiness and this is probably seldom the case. Because of the difficulties of defining happiness it is also difficult to define how one can achieve it. High consumption possibility is however equal to the possibility to consume at a high level. This means that it would be possible to get food, housing, healthcare, education et cetera and this is more likely to bring welfare than well-being.

## 2.4 Other indexes on welfare and well-being

Human Development Index (HDI), Democratic Index (DI), Happy Planet Index (HPI), Index of Sustainable Economic Welfare (ISEW), Genuine Progress Indicator (GPI), Sustainable Net Benefit Index (SNBI), Physical Quality of Life Index (PQLI), Legatum Prosperity Index, Gross National Happiness (GNH) et cetera are all indexes that includes different variables in order to define what they argue is welfare and well-being. The text which follows will describe each and every one of these indexes with the focus on the HDI and DI because these two will be compared with the Generated Index of Welfare in this essay.

### 2.4.1 HDI – Human Development Index

"The HDI – human development index – is a summary composite index that measures a country's average achievements in three basic aspects of human development: health, knowledge, and a decent standard of living. Health is measured by life expectancy at birth; knowledge is measured by a combination of the adult literacy rate and the combined primary, secondary, and tertiary gross enrolment ratio; and standard of living by GDP per capita (PPP US$)."(UNDP 2010)

The HDI aims to capture relevant aspects of human well-being. (Kulig, Kolfoort et al. 2010) The HDI is calculated in the following way, see Figure 1 (UNDP 2008).

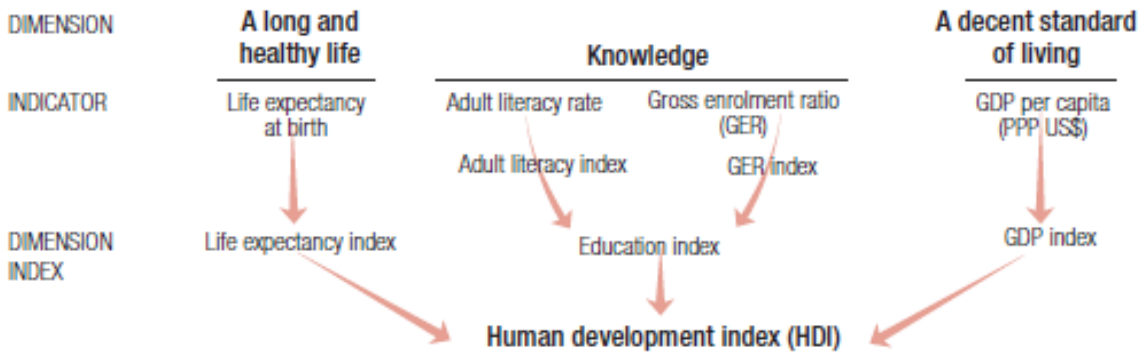



**Figure 1. Calculating HDI, (UNDP 2008)**

HDI is widely used around the world and is often taken as the single most important factor along with GDP to measure a countries success and well-being. The fact that this more or less ad-hoc index is so widely used can be problematic seeing that there is a large amount of arbitrary in the index.

The problems surrounding HDI is widely discussed and the index has been criticized on several occasions during the years of its existence. For example the index has not taken any kind of ecological indicators in consideration which flaws the index in a global perspective. Even if the intention of HDI is and was good much have happened since its beginning and (Sagara and Najam 1998) argues that the human development reports "have lost touch with their original vision and the index fails to capture the essence of the world it seeks to portray".

A humorous but in the same way interesting point of view is made by the economist Bryan Caplan:

“Now what exactly is the HDI? The one-line explanation is that it gives "equal weights" to GDP per capita, life expectancy, and education. But it's more complicated than that, because scores on each of the three measures are bounded between 0 and 1. This effectively means that a country of immortals with infinite per-capita GDP would get a score of .666 (lower than South Africa and Tajikistan) if its population were illiterate and never went to school.” (Caplan 2009)

Caplan (2009) also addresses the problems with how HDI is weighted between the different variables and argues that GDP per capita and life expectancy should be more important than education. Further he argues that to be able to get full marks on the education score of HDI 100% of the population needs to be students. (Caplan 2009)

The index can also be criticized from different points of view which Ratan La Basu (2005) does. Instead of only focusing on the gain or loss of material possessions Ratan La Basu (2005) argues that the spiritual and moral development needs to be taken in consideration to improve the index. (Basu 2005)

### 2.4.2 DI – Democracy Index

The weekly news and international affairs publication *The Economist* have created an index called the Democracy Index. The focus of this index is based on five general categories: electoral process and pluralism, civil liberties, functioning of government, political participation and political culture. 167 countries have been classified with a democracy index as of 2008 and the scale of the index is from zero to ten and depending on the countries score they are placed into one of four categories: "Full Democracies" (score 8-10), "Flawed Democracies" (score 6-7.9, "Hybrid Regimes" (all considered democracies) (score 4-5.9), and "Authoritarian Regimes" (considered dictatorial) (score below 4).

The DI is a weighted score from the answer of 60 questions each with two or three alternatives. The 60 questions are split into five categories and each answer is translated to a mark of either 0 or 1 or 0.5 for the three answer alternative questions. The sums are then added in each category, multiplied by ten and divided by the total number of questions within the category. If one question, e.g. “if the elections for national legislation and head of government are not considered free” results in a mark of 0 then the following question "Are elections... fair?" is not taken in consideration and automatically yields a 0. (Economist 2008)

Some questions are considered too important that if they get a low score they give a penalty on the total score sum in their different categories, and they are:

1. "Whether national elections are free and fair";
2. "The security of voters";
3. "The influence of foreign powers on government";
4. "The capability of the civil servants to implement policies".

(Economist 2008)

The main problem with democracy index is the short time period of which it has been in use. It has been created two times with the first published 2006 and the second in 2008 and it needs several years of analyzing in order to find out if this index has the possibility to determine democracy or not. Because of the procedure of the index involves questionnaires the index can't be created for years before 2006 which is problematic in order to make conclusions over time. Most other indexes use available statistics and can be calculated as long as the sufficient statistics is collected.

However, even if the HDI could be calculated for years before 1975 there is no record of it and the UNDP argues that only life expectancy is available in 1950 (UNDP 2010).

### 2.4.3 HPI – Happy Planet Index

The New Economics Foundation (NEF) introduced the Happy Planet Index in July 2006 as an index of human well-being and environmental impact.

HPI incorporates three separate indicators: ecological footprint, self-reported life satisfaction, and life expectancy. The formula for the HPI is, Equation 7:

$$HPI = \frac{Life\ Satisfaction * Life\ Expectancy}{Ecological\ Footprint} \qquad \textbf{Equation 7}$$

The index aims to reflect the average years of happy life produced by a given society, nation or group of nations, per unit of planetary resources consumed. In other words, the efficiency which countries convert the earth's finite resources into well-being experienced by their citizens. (NewEconomicsFoundation 2006)

The problem with HPI, as with several other indexes, is that the "life satisfaction" in this case is a very subjective matter and difficult to measure. If one of the main components of an index is very subjective the index itself becomes subjective and the quality of it lowers immensely.

### 2.4.4 ISEW - Index of Sustainable Economic Welfare

The ISEW is a type of indicator intended to measure sustainability and economic welfare without the limitations of GDP by including the value of externalities, income distribution and natural resources depletion. (Daly and Cobb 1989) The GDP does not account for the way income is distributed across society and includes expenses needed to maintain the well-being level, such

as those resulting from the environmental and social conditions generated by market economy itself. (Beça and Santos 2010)

Daly and Cobb calculated the ISEW for the USA and have since been calculated for most western European and Scandinavian countries, Canada, UK, Australia and Chile. The methods used to calculate the index have been modified and revised during the years. Different names have been given for the ISEW such as a Genuine Progress Indicator (GPI) and a Sustainable Net Benefit Index (SNBI), there are however differences which will be clarified below. The macroeconomic system is beneficial to human well-being up to a point but beyond this the growth appears to be damaging. Initially the three indexes supports the need for countries to focus on sustainable qualitative improvement (sustainable development, SD) instead of growth. (Lawn 2003)

The main idea of the ISEW and GPI is to more close approximate the sustainable economic welfare or progress of a nation's citizens. The sustainable economic welfare is in this case the welfare at a particular point in time while taking the past and present activities into consideration. The SNBI differs from GPI and ISEW where the included variables are sorted into separate 'benefit' and 'cost' accounts. The total of the cost account is subtracted from the benefit account to obtain the SNBI. The advantage of this is that the results can be presented with an income and capital concept superior to standard definitions of income. The possibility of comparing the benefits and costs of a growing macro economy is another advantage. (Lawn 2003)

While the ISEW criticize GNP (gross national product) for its deficiencies as a welfare indicator the ISEW can be criticized on the fact that GNP never was thought of as a welfare indicator in the first place. The ISEW is highly dependent on certain key and rather arbitrary assumptions about the weighting of income distribution, the neglect of technical progress and increases in human capital and the valuing of the depletion of non-renewable resources and long-term environmental damage. The theoretical foundation of the ISEW is accused of being flawed and it rests on two methodological inconsistencies where the first is that ISEW lumps together the measurement of current welfare and sustainability which can be argued to be kept apart. The second part is that although ISEWs are usually undertaken by economists, in favor of some stronger version of sustainability, the ISEW implicitly assumes perfect substitutability within natural capital and between natural and other forms of capital. (Neumayer 1999)

### 2.4.5 PQLI – Physical Quality of Life Index

This index attempts to create a practical measure of social distribution that will avoid the limitations of GNP, minimize cultural and developmental ethnocentricity and be internationally comparable. PQLI uses three indicators: literacy, infant mortality, and life expectancy at age one. Each indicator is placed on a fixed scale between 1 and 100 where the worst (=1) is set on the basis of historical experience and the best (=100) is fixed in terms of what might be achieve within the next half-century or so. The three indicators are placed on a common scale to create a composite index of an equally weighted summation. The PQLI aims to measure result, not inputs. (Morris 1977)

The problem with PQLI is that it lacks comparability across countries of the underlying data; because of the literacy rates used in the calculation of the PQLI were taken from years ranging from 1946 to 1973. Instead of reflecting a fundamental disparity in level of development the reported PQLIs only show a lack of comparability of the data. (Brodsky and Rodrik 1981)

### 2.4.6 Legatum Prosperity Index

The Legatum Prosperity Index of 2009 is based on 79 different variables on 104 nations around the world. 9 sub-indexes are grouped from the 79 different variables:

1. Economic Fundamentals
2. Democratic Institutions
3. Health
4. Governance
5. Social Capital
6. Entrepreneurship and Innovation
7. Education
8. Safety and Security
9. Personal Freedom

(LegatumInstitute 2009)

There are still barely any scientific articles on the Legatum Prosperity Index and more research should be undertaken in order to evaluate its properties both the positive and negative. A search in Electronic Library Information Navigator (ELIN@Lund) results in merely two hits on ”Legatum Prosperity Index”, none of which from any journal of scientific relevance.

### 2.4.7 GNH – Gross National Happiness

As well as other indicators (as the ISEW for example) the gross national happiness was developed to present a quality of life in other more psychological and holistic terms than gross national product or GDP. The 4th King of Bhutan HM Jigme Singye Wangchuck, coined the term GNH and promulgated it in the beginning of his reign in 1972.

The index comprise of nine normative and equally weighted grounds and within each of this dimensions several indicators was chosen that seemed likely to remain informative across time, were relatively uncorrelated and had high response rates. The nine dimensions are:

1. Psychological Well-being
2. Time Use
3. Community Vitality
4. Culture
5. Health
6. Education
7. Environmental Diversity
8. Living Standard
9. Governance

(Ura 2008)

As well as other indexes GNH depends on a series of subjective judgments about well-being. This lets governments define the GNH in a way that suites their own interests and the possibility of comparison is more or less depleted.

It has been suggested that better measures of the subjective well-being can be created by extensive use of complex questionnaires that seek to measure psychological well-being and satisfaction with various aspects of life including material wellbeing, health and relationships. (Winton 2009)

However there is always problems connected with questionnaires of any kind, and if the data cannot be measured in a clear and distinct way it can never be close to absolute.

### 2.4.8 Is some index better than others and are they useful?

Overall the described indexes have several problems with some highs and lows. The HDI has a far to firm grip in the welfare discussions and is in need of wider perspective with both more and different included variables. The DI along with the Legatum Prosperity Index needs more research because in the present they both lack a wide enough material of discussion. The fact that the ISEW lumps together welfare and sustainability is not necessary a bad thing even if Neumayer (1999) think it is because these two approaches broadens the point of view and includes the environment in a way that several other indexes lacks. HPI on the other hand focus too much on the environment and lacks other parts and evade concepts like trade and innovation. The lack of these parts is also true for the GNH which above all is too subjective to be of any interest in this study. Some indexes like the PQLI have comparability problems which is a severe flaw that cripples the intention of the index.

## 2.5 Variables of choice and definitions needed hereby

Some definitions are needed in order to figure out which variables to be included in the study.

When choosing the variables of interest the focus on welfare will be in the well-being direction which is a subjective measure of the satisfaction or happiness of individuals. The main reason behind this is to make the constructed index comparable with the HDI which aims to capture relevant aspects of human well-being. (Kulig, Kolfoort et al. 2010)

The chosen variables need to be diverse to capture both values important to the developed world as well as the developing world. The economic variables are the simplest to choose because they are almost equally important to all of the countries. The environmental variables are a bit harder because for example a high emission of carbon dioxide usually is equal to a highly developed industry but the well-being of the citizens may be lowered if the climate declines because of the emissions. The socio-economic variables are probably the most difficult to choose because of the wide spread of countries. In Sweden the amount of paved roads are not going to differ a lot during the time period but for a country like Sri Lanka they will show to increase heavily in the late 1990s. The choice of variables is of course one of the biggest sources of error in the study, but on the bright side this can relatively easily be changed if the model itself is proven to be adequate.

### 2.5.1 Do free markets foster political freedoms?

In order to compare the Generated Index of welfare with democracy index the variables needs to be chosen in a way that describes democracy. The following theory describes the connection between trade and democracy which makes it possible to choose trade variables which describes democracy.

"Economic openness and the commercial competition and contact it brings can directly and indirectly promote civil and political freedoms within countries." (Griswold 2004)

Griswold argues that when a country trade with another country it will have an increasing contact between the nation's citizens experience with the rest of the world including all types of contact platforms such as telephone, internet and face-to-face contact. When this portal has been opened ideas and new ways of thinking can be shared between the countries in question together with both industrial and consumer goods in form of magazines, books, digital media which in themselves carry both social and political content. This connectivity with foreign investment generates the possibility for foreign study and travel to experience other nations with their cultures. (Griswold 2004)

"Theory and evidence together argue that trade liberalization and a more general openness to the global economy do correlate with more political and civil freedom, in the world as a whole and within individual countries." (Griswold 2004)

# 3 Method

## 3.1 Principal Component Analysis (PCA)

Principal component analysis is a multivariate technique that analyzes a data table in which observations are described by several inter-correlated quantitative dependent variables. The goals of PCA are to find and extract the most important information from the data and compress the size while at the same time keeping the important information and simplify the description of data, and then the structure of the observations and variables can be analyzed. (Abdi and Williams 2010)

The PCA computes new variables called principal components (PCs) as linear combinations of the original variables. The first principal component is required to have the largest possible variance (in other words inertia and therefore explain the largest part of the inertia of the data table). The second has to be orthogonal to the first and have the second largest possible inertia. The rest of the components are computed likewise. The values of these new variables for the observations are called factor scores, which can be interpreted geometrically as the projections of the observations onto the principal components. (Abdi and Williams 2010)

To be able to find the principal components there is a need for both vectors and matrixes and for those interested in the mathematics behind this be sure to study the Appendix at paragraph A2 Principal Component Analysis: finding the components at page 59.

The strengths of the principal component analysis are that a large amount of variables can be used without adding much to the complexity of the model.

## 3.2 Standardization of data and test for similarities

In order to compare the data there is naturally a need for standardization and the method of doing this is to be found in the Appendix at A3 Standardization of Data at page 60. Theory for linear regression with residual analysis in form of normal distribution and test for Heteroscedasticity is also found in the Appendix, A5 Linear regression at page 63-65.

## 3.3 Creating an index

From the chosen groups of variables which is economic, socio-economic and environment a new index is to be created. From the principal component analysis the scores and percent of variance explained are given for all the countries and variables. This can be done with the following formula, Equation 8, (Antony and Rao 2007):

$$H_j = \sum_{i=1}^{14} s_i * p_i \qquad \textbf{Equation 8}$$
$$, where\ j = 1,2 \dots 115$$

The score for factor $i$ $(s_i)$ is multiplied with the percent explained for that factor $(p_i)$ and so on until all the scores and factors for that country has been summarized$(H_j)$. This is then repeated for all the countries until all 115 countries have an index. Then the index is standardized to a scale from 0 to 100 with the following equation, Equation 9, (Antony and Rao 2007):

$$I_j = \frac{H_j - H_{\text{minimum}}}{H_{maximum} - H_{minimum}} \qquad \textbf{Equation 9}$$
$$, where\ j = 1,2 \dots 115$$

Last the numbers are sorted to get an order from 1 to 115 and the index have thus been created.

The unscaled index is also used in some parts of this essay and is then specified with "Unscaled" in from of the Generated Index then the index is not standardized to a scale, see paragraph 5.5 Time series analysis at page 49.

This essay will use a time span of ten years when calculating the Generated Index of Welfare as a mean value of 1998-2008 in order to be able to compare the index with HDI taken from the same period and DI which only exists for a pair of years. This will of course affect the index in several ways. The economy has fluctuated greatly in the past years which could make the generated weights for the countries far to unbalance to be used over time. But the economy has suffered several crises from the 1980s to present and because a comparison with the HDI is desirable it is difficult to find other obvious candidates for the time span. The countries will only be compared with the included 115 countries which can affect the quality of the index and it

would of course be better if all countries could have been included, but due to the lack of statistics this is not possible.

## 3.4 Verifying the results

In order to verify the results of the principal component analysis several different approaches can be made. In this essay the biplot and dendrogram have been chosen. A biplot is a form of graphical representation of the principal component analysis which can help in the interpretation of the PCA. It will among other things provide information of how the variables are connected to each other. The dendrogram, on the other hand, is a cluster analysis model which groups together similarities and should provide information if the Generated Indexes on welfare are correctly distributed.

### 3.4.1 Biplot

The information computed in the principal component analysis can be displayed graphically with both samples and variables in the same plot. A two-dimensional plot can be useful for detecting patterns in the data. The variables are displayed as vectors and the samples are displayed as points. If the angle between a point and a vector is small (<<90°) this implies a positive correlation, the opposite with large angles (>>90°) implies a negative correlation. The length of the vectors and points from origo explain the strength of the correlation. (Gower and Hand 1996) and (Adler, Yazhemsky et al. 2010)

Figure 2 below depict the difference between the two extremes. The yellow circle on the right is at the same length as the thick red vector tip and at a small angle which make their correlation positive and strong. The green square on the left is at a shorter length from origo than the thick red vector with a large angle which make their correlation negative and weak. The two black arrows are not vectors but only in the figure to clarify the distance from origo and angle between the circle/square and the vector.

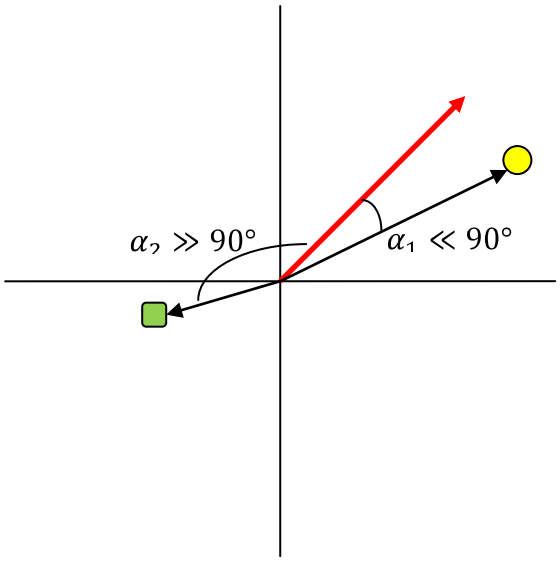



**Figure 2. Example of a Biplot**

### 3.4.2 Cluster analysis and dendrogram

In order to verify the principal component analysis a form of cluster analysis is used, called dendrogram. Cluster analysis is a type of unsupervised pattern recognition and was originally developed by biologist to find similarities between organisms. In order to determine the relationships between different species the numerical taxonomy emerged. The most common form of cluster analysis knows to the general public is probably the usage of dendrogram to form family trees. This is an intuitive simple method of understanding correlation between different groups (in this essay the countries). More about the theory behind dendrogram can be found in the Appendix at page 58. (Brereton 2003)

## 3.5 Hypothesis

A hypothesis is formed to be used as a guideline to the empirical study.

The ISEW share several interesting aims with this essay however it is not this essay intention to remove the GDP but to include it (GDP growth, in this case) together with other economic factors, socio-economic factors and environmental factors. By combing the parts an overall index is intended to be created.

It is plausible to assume that similar developed countries will connect in clusters in a relatively clear way. It is assumed that the developed countries will have high ranking and be mainly be the Western World including Europe, Australia, Canada, USA.

In short this breaks down into:

Industrially developed and democratic countries will have a high Generated Index of Welfare (GI).

# 4 Empirical Analysis

## 4.1 Collection of data

The gathered data for the 14 factors and GNI per capita were collected from the World Development Indicators Online (WDI) which is the World Bank's premier annual compilation of data about development and the global economy. (TheWorldBank 2010)

The data for the HDI were collected from the United Nations Development Programme (UNDP) which is the UN's global development network. (UNDP 2010)

The data for the DI were collected from the Economist, The Economist Intelligence Unit's Index of Democracy 2008. (Economist 2008)

## 4.2 Validation & trustworthiness

The data from the World Bank and UNDP are both from well known institutes over a long series of time and are therefore assumed to be trustworthy. The data from The Economist is only for a specific year which makes the comparison blunter and this is one of the reasons why only a time span of 10 years have been chosen from the 14 studied factors. The Economist itself is a well known weekly news and international affairs publication dating from 1843 and is also assumed to be trustworthy.

### 4.2.1 Over time

The main problem when comparing several countries over time is the lack of statistics. In case of missing statistics the subsequent year (or the previous if there is now subsequent) has been used as the present. This will of course foster the source of error which needs to be taken in consideration while interpreting the results.

## 4.3 Source criticism

The main criticism on the chosen data can be focused on two part. The first is the one described above about the missing statistics and the second is the short time span of the Democratic Index. But without the comparison with the Democratic Index the opinion of that the Generated Index also should be a measure of the trade and democratic situation in a country

seems irresponsible. Both the World Bank and UNDP have statistics over a fairly long time span and are both international well known institutes.

## 4.4 Data

In order to compare several countries to find an index on welfare factors from mainly three different categories have been chosen: economic, socio-economic and environmental. The factors need to be chosen in a way that represents the welfare and the trade of the country. The reason behind this is that a comparison between the Generated Index, HDI and DI is sought after. 115 countries with 14 factors have been studied.

The fourteen factors are as follows: GDP growth (annual %), Inflation. GDP deflator (annual %), High-technology exports (% of manufactured exports), Immunization. measles (% of children ages 12-23 months), Internet users (per 100 people), Life expectancy at birth. total (years), Mobile cellular subscriptions (per 100 people), Population growth (annual %), Roads. paved (% of total roads), Mortality rate. under-5 (per 1.000), Time required to start a business (days), Renewable internal freshwater resources per capita (cubic meters), Electric power consumption (kWh per capita), CO2 emissions (metric tons per capita). Figure 3 below depict some of the chosen standardized factors for Sweden plotted over time (1960-2008).Lists of included countries can be found in paragraph A7 Countries at page 71 in the Appendix.

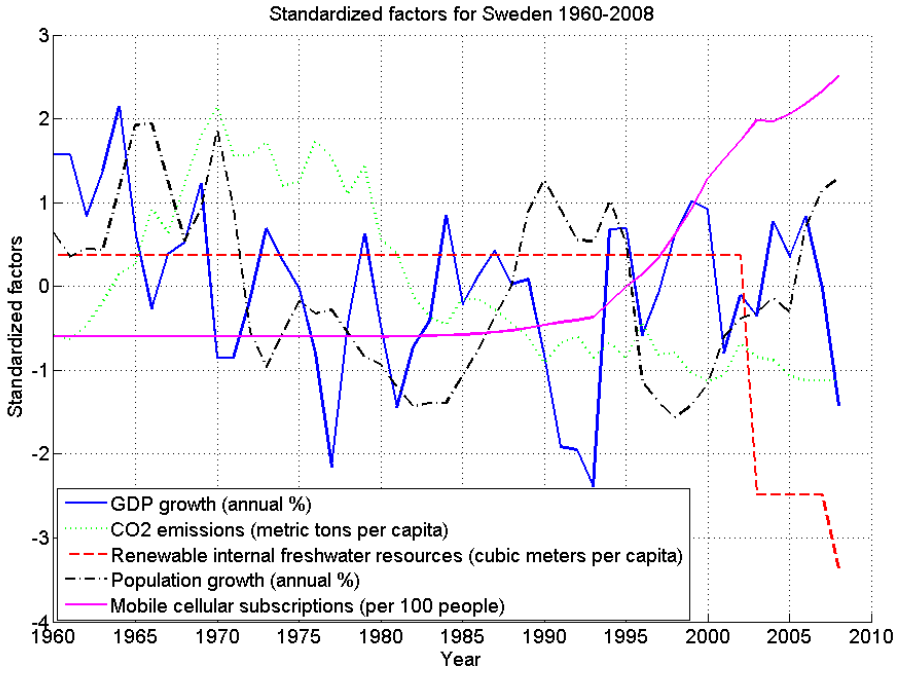



**Figure 3. Some of the chosen standardized factors are plotted for Sweden over time**

## 4.5 Simulation tool

The computer software tool MATLAB is used in the analysis and stands for "MATrix LABoratory" which is a numerical computing environment and fourth-generation programming language. It has the possibility to perform principal component analysis as well as cluster analysis in a lucid way. To the untrained eye it differs drastically from the program Microsoft Excel but in fact they share several common functions and possibilities.

The programmed MATLAB code for this essay spans over thirty pages and can thus be supplied upon request.

# 5 Analysis

## 5.1 Generated Index of Welfare (GI)

A new index has been created for the included countries and is shown in the table below. Both the values and the ranks for the GI as well as the HDI, DI and GNI per capita (PPP) are included in the table, Table 1. GNI per capita (PPP) has been standardized in the same way is the GI as described in Equation 9 at page 19. In Table 1 all the 14 factors have been included to make a Generated Index with 100 % of the variation explained. Later in the essay this number is reduced in order to see if fewer components also make out a stable model, at 5.3 Test of stability.

Iceland finishes first and is, as described later, the country that differ the most in a positive way as Mozambique is the complete opposite. Some of the largest economies in the world places in the top such as the United States and Japan and several European countries is in the top 30. The first African country on the list is Egypt at position 70.

As expected mostly well developed countries such as Norway, Canada, Finland, the United States, Sweden and Denmark tops the list of the Generated Index of Welfare and in the bottom we find developing countries such as Namibia, Cameroon, Ethiopia, Nigeria, Cambodia and Mozambique.

The correlations of the different values in Table 1 are analyzed in 5.4.3 at 44.

**Table 1. Generated Index with 14 factors included together with HDI, DI and GNI per capita (PPP) values and rank**

| Countries | 14 factors (All) Value | HDI Value | DI Value | GNI per capita (PPP) Value | 14 factors All Rank | HDI Rank | DI Rank | GNI per capita (PPP) Rank |
|---|---|---|---|---|---|---|---|---|
| Iceland | 100.0 | 96.1 | 96.5 | 58.2 | 1 | 3 | 3 | 15 |
| Norway | 79.2 | 96.8 | 96.8 | 81.1 | 2 | 1 | 2 | 4 |
| Luxembourg | 70.0 | 51.7 | 91.0 | 100.0 | 3 | 104 | 9 | 1 |
| Canada | 69.7 | 96.1 | 90.7 | 60.7 | 4 | 4 | 11 | 10 |
| Finland | 69.3 | 95.1 | 92.5 | 55.6 | 5 | 12 | 6 | 20 |
| United States | 68.5 | 95.4 | 82.2 | 75.6 | 6 | 11 | 18 | 5 |
| United Arab Emirates | 67.4 | 88.6 | 26.0 | 83.8 | 7 | 30 | 147 | 3 |
| Sweden | 66.8 | 96.0 | 98.8 | 59.3 | 8 | 5 | 1 | 12 |
| Singapore | 64.4 | 94.3 | 58.9 | 72.6 | 9 | 20 | 82 | 6 |
| Denmark | 64.0 | 94.9 | 95.2 | 60.7 | 10 | 15 | 5 | 9 |
| Australia | 63.6 | 96.5 | 90.9 | 58.7 | 11 | 2 | 10 | 14 |
| Netherlands | 63.3 | 95.8 | 95.3 | 64.3 | 12 | 6 | 4 | 8 |
| New Zealand | 61.5 | 94.4 | 91.9 | 41.8 | 13 | 19 | 7 | 27 |
| Switzerland | 60.9 | 95.6 | 91.5 | 70.9 | 14 | 8 | 8 | 7 |
| United Kingdom | 60.7 | 94.3 | 81.5 | 58.2 | 15 | 21 | 21 | 16 |
| Japan | 59.3 | 95.4 | 82.5 | 56.0 | 16 | 9 | 17 | 19 |
| Germany | 59.2 | 94.5 | 88.2 | 56.6 | 17 | 17 | 13 | 18 |
| Korea, Rep. | 58.4 | 91.7 | 80.1 | 39.4 | 18 | 24 | 28 | 30 |
| Austria | 57.4 | 94.9 | 84.9 | 60.1 | 19 | 13 | 14 | 11 |
| Italy | 57.3 | 94.4 | 79.8 | 52.4 | 20 | 18 | 29 | 22 |
| Bahrain | 56.5 | 88.5 | 33.8 | 46.8 | 21 | 31 | 130 | 24 |
| Belgium | 55.9 | 94.9 | 81.6 | 58.8 | 22 | 14 | 20 | 13 |
| Ireland | 55.4 | 95.7 | 90.1 | 56.8 | 23 | 7 | 12 | 17 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **France** | 55.2 | 95.4 | 80.7 | 54.4 | 24 | 10 | 24 | 21 |
| **Czech Republic** | 53.1 | 89.1 | 81.9 | 33.7 | 25 | 29 | 19 | 34 |
| **Israel** | 52.8 | 92.6 | 74.8 | 43.3 | 26 | 23 | 38 | 25 |
| **Portugal** | 49.2 | 90.4 | 80.5 | 35.6 | 27 | 28 | 25 | 32 |
| **Slovenia** | 49.0 | 91.6 | 79.6 | 40.1 | 28 | 25 | 30 | 29 |
| **Malaysia** | 48.1 | 81.8 | 63.6 | 40.4 | 29 | 50 | 68 | 28 |
| **Cyprus** | 48.1 | 90.8 | 77.0 | 18.8 | 30 | 27 | 36 | 46 |
| **Estonia** | 47.8 | 86.7 | 76.8 | 25.1 | 31 | 35 | 37 | 38 |
| **Greece** | 47.7 | 92.8 | 81.3 | 42.9 | 32 | 22 | 22 | 26 |
| **Spain** | 47.5 | 94.7 | 84.5 | 47.6 | 33 | 16 | 15 | 23 |
| **Slovak Republic** | 45.4 | 86.5 | 73.3 | 27.0 | 34 | 36 | 44 | 37 |
| **Croatia** | 45.1 | 85.9 | 70.4 | 24.5 | 35 | 38 | 51 | 39 |
| **Hungary** | 43.7 | 86.9 | 74.4 | 27.0 | 36 | 33 | 40 | 36 |
| **Jamaica** | 42.4 | 76.2 | 72.1 | 11.1 | 37 | 70 | 49 | 65 |
| **Poland** | 42.1 | 87.0 | 73.0 | 23.3 | 38 | 32 | 45 | 40 |
| **Latvia** | 41.4 | 60.1 | 72.3 | 20.3 | 39 | 95 | 46 | 43 |
| **Lithuania** | 41.3 | 80.1 | 73.6 | 21.9 | 40 | 58 | 42 | 41 |
| **Russian Federation** | 41.1 | 81.1 | 44.8 | 18.7 | 41 | 52 | 107 | 47 |
| **Brunei Darussalam** | 40.9 | 91.5 | No data | 88.7 | 42 | 26 | No data | 2 |
| **Bulgaria** | 40.8 | 82.7 | 70.2 | 14.5 | 43 | 46 | 52 | 58 |
| **Chile** | 39.8 | 86.8 | 78.9 | 19.1 | 44 | 34 | 32 | 45 |
| **Macedonia, FYR** | 39.2 | 81.0 | 62.1 | 37.7 | 45 | 53 | 72 | 62 |
| **Saudi Arabia** | 39.2 | 84.0 | 19.0 | 12.2 | 46 | 44 | 161 | 31 |
| **Trinidad and Tobago** | 38.1 | 82.5 | 72.1 | 29.7 | 47 | 48 | 48 | 35 |
| **Thailand** | 38.0 | 77.3 | 68.1 | 10.3 | 48 | 65 | 54 | 67 |
| **Uruguay** | 37.9 | 85.4 | 80.8 | 16.6 | 49 | 39 | 23 | 52 |
| **Ukraine** | 37.7 | 78.1 | 69.4 | 8.2 | 50 | 63 | 53 | 76 |
| **Oman** | 36.4 | 84.2 | 29.8 | 33.8 | 51 | 43 | 140 | 33 |
| **Romania** | 36.3 | 82.0 | 70.6 | 14.9 | 52 | 49 | 50 | 56 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Kazakhstan** | 35.5 | 78.6 | 34.5 | 12.0 | 53 | 62 | 127 | 63 |
| **Argentina** | 35.1 | 86.1 | 66.3 | 18.4 | 54 | 37 | 56 | 49 |
| **Jordan** | 34.9 | 74.8 | 39.3 | 6.8 | 55 | 74 | 117 | 78 |
| **Turkey** | 34.3 | 79.1 | 56.9 | 18.2 | 56 | 61 | 87 | 50 |
| **Tunisia** | 34.1 | 74.2 | 29.6 | 9.9 | 57 | 77 | 141 | 71 |
| **Panama** | 34.0 | 82.9 | 73.5 | 15.0 | 58 | 45 | 43 | 55 |
| **Mexico** | 33.8 | 84.3 | 67.8 | 20.3 | 59 | 42 | 55 | 44 |
| **Moldova** | 32.8 | 70.8 | 65.0 | 3.2 | 60 | 86 | 62 | 93 |
| **Lebanon** | 32.2 | 84.7 | 56.2 | 16.1 | 61 | 40 | 89 | 53 |
| **Iran, Islamic Rep.** | 31.8 | 76.8 | 28.3 | 15.2 | 62 | 67 | 145 | 54 |
| **Kyrgyz Republic** | 31.7 | 70.1 | 40.5 | 5.1 | 63 | 87 | 114 | 100 |
| **Morocco** | 31.7 | 63.1 | 38.8 | 1.9 | 64 | 94 | 120 | 87 |
| **Belarus** | 31.2 | 81.1 | 33.4 | 13.4 | 65 | 51 | 132 | 61 |
| **Syrian Arab Republic** | 31.2 | 73.2 | 21.8 | 6.1 | 66 | 80 | 156 | 82 |
| **Bosnia and Herzegovina** | 30.7 | 80.7 | 57.0 | 10.1 | 67 | 54 | 86 | 68 |
| **Algeria** | 30.4 | 74.1 | 33.2 | 11.0 | 68 | 78 | 133 | 66 |
| **Georgia** | 30.2 | 76.3 | 46.2 | 5.0 | 69 | 69 | 104 | 88 |
| **Egypt, Arab Rep.** | 30.0 | 69.1 | 38.9 | 11.7 | 70 | 88 | 119 | 64 |
| **Colombia** | 30.0 | 79.4 | 65.4 | 7.0 | 71 | 59 | 60 | 77 |
| **Costa Rica** | 29.8 | 84.3 | 80.4 | 14.7 | 72 | 41 | 27 | 57 |
| **South Africa** | 29.0 | 68.2 | 79.1 | 13.9 | 73 | 91 | 31 | 60 |
| **Albania** | 28.9 | 80.7 | 59.1 | 9.2 | 74 | 55 | 81 | 73 |
| **Armenia** | 28.4 | 77.5 | 40.9 | 5.8 | 75 | 64 | 113 | 83 |
| **China** | 28.3 | 75.3 | 30.4 | 5.8 | 76 | 71 | 136 | 84 |
| **Philippines** | 28.2 | 74.2 | 61.2 | 4.6 | 77 | 76 | 77 | 90 |
| **Paraguay** | 27.6 | 75.2 | 64.0 | 6.4 | 78 | 72 | 66 | 80 |
| **Sri Lanka** | 27.2 | 74.9 | 66.1 | 5.2 | 79 | 73 | 57 | 86 |
| **Guatemala** | 26.9 | 68.9 | 60.7 | 6.5 | 80 | 89 | 79 | 79 |
| **El Salvador** | 26.8 | 73.5 | 64.0 | 9.2 | 81 | 79 | 67 | 74 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Dominican Republic** | 26.4 | 76.5 | 62.0 | 9.8 | 82 | 68 | 73 | 72 |
| **Ecuador** | 25.8 | 80.6 | 56.4 | 10.1 | 83 | 56 | 88 | 69 |
| **Mongolia** | 25.3 | 70.9 | 66.0 | 3.5 | 84 | 85 | 58 | 92 |
| **Gabon** | 25.2 | 74.7 | 30.0 | 20.7 | 85 | 75 | 139 | 42 |
| **Nicaragua** | 24.9 | 68.8 | 60.7 | 2.9 | 86 | 90 | 78 | 94 |
| **Honduras** | 24.3 | 71.9 | 61.8 | 4.6 | 87 | 82 | 74 | 89 |
| **Peru** | 24.1 | 79.2 | 63.1 | 9.9 | 88 | 60 | 70 | 70 |
| **Venezuela, RB** | 23.2 | 82.5 | 53.4 | 17.3 | 89 | 47 | 95 | 51 |
| **Vietnam** | 22.8 | 71.3 | 25.3 | 2.5 | 90 | 84 | 149 | 98 |
| **Bolivia** | 21.5 | 71.9 | 61.5 | 5.6 | 91 | 81 | 75 | 85 |
| **Pakistan** | 21.4 | 56.5 | 44.6 | 2.9 | 92 | 98 | 108 | 95 |
| **Brazil** | 19.9 | 80.4 | 73.8 | 14.2 | 93 | 57 | 41 | 59 |
| **Nepal** | 19.5 | 53.4 | 40.5 | 0.7 | 94 | 100 | 115 | 112 |
| **Azerbaijan** | 18.8 | 77.2 | 31.9 | 6.1 | 95 | 66 | 135 | 81 |
| **Indonesia** | 17.7 | 71.5 | 63.4 | 4.1 | 96 | 83 | 69 | 91 |
| **Cote d'Ivoire** | 16.7 | 48.2 | 32.7 | 1.8 | 97 | 110 | 134 | 101 |
| **Togo** | 16.3 | 49.7 | 24.3 | 0.4 | 98 | 108 | 151 | 113 |
| **India** | 15.3 | 59.2 | 78.0 | 2.8 | 99 | 96 | 35 | 96 |
| **Benin** | 14.3 | 47.7 | 60.6 | 1.3 | 100 | 111 | 80 | 106 |
| **Yemen, Rep.** | 14.1 | 55.7 | 29.5 | 2.7 | 101 | 99 | 142 | 97 |
| **Bangladesh** | 13.9 | 52.5 | 55.2 | 0.9 | 102 | 102 | 91 | 109 |
| **Senegal** | 13.8 | 45.6 | 53.7 | 18.5 | 103 | 113 | 93 | 48 |
| **Botswana** | 13.8 | 67.1 | 74.7 | 1.8 | 104 | 93 | 39 | 102 |
| **Sudan** | 13.4 | 51.6 | 28.1 | 1.6 | 105 | 105 | 146 | 104 |
| **Kenya** | 13.3 | 53.2 | 47.9 | 1.4 | 106 | 101 | 103 | 105 |
| **Tanzania** | 13.0 | 50.4 | 52.8 | 0.8 | 107 | 107 | 96 | 111 |
| **Ghana** | 12.5 | 51.3 | 53.5 | 1.0 | 108 | 106 | 94 | 108 |
| **Zambia** | 12.1 | 46.3 | 52.5 | 0.8 | 109 | 112 | 97 | 110 |
| **Namibia** | 11.3 | 67.4 | 64.8 | 8.6 | 110 | 92 | 64 | 75 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Cameroon** | 10.9 | 51.9 | 34.6 | 2.4 | 111 | 103 | 126 | 99 |
| **Ethiopia** | 10.2 | 38.5 | 45.2 | 0.0 | 112 | 115 | 105 | 114 |
| **Nigeria** | 7.6 | 49.6 | 35.3 | 1.7 | 113 | 109 | 124 | 103 |
| **Cambodia** | 5.7 | 56.7 | 48.7 | 1.3 | 114 | 97 | 102 | 107 |
| **Mozambique** | 0.0 | 38.5 | 54.9 | 0.0 | 115 | 114 | 92 | 115 |

## 5.2 Scree plot from the principal component analysis (PCA)

With the use of the information gathered from the principal component analysis (eigenvalues of the covariance matrix), in paragraph 3.1 at page 18, a scree plot can be created depicting the fraction of total variance in the data as explained or represented by each principal component, Figure 4.

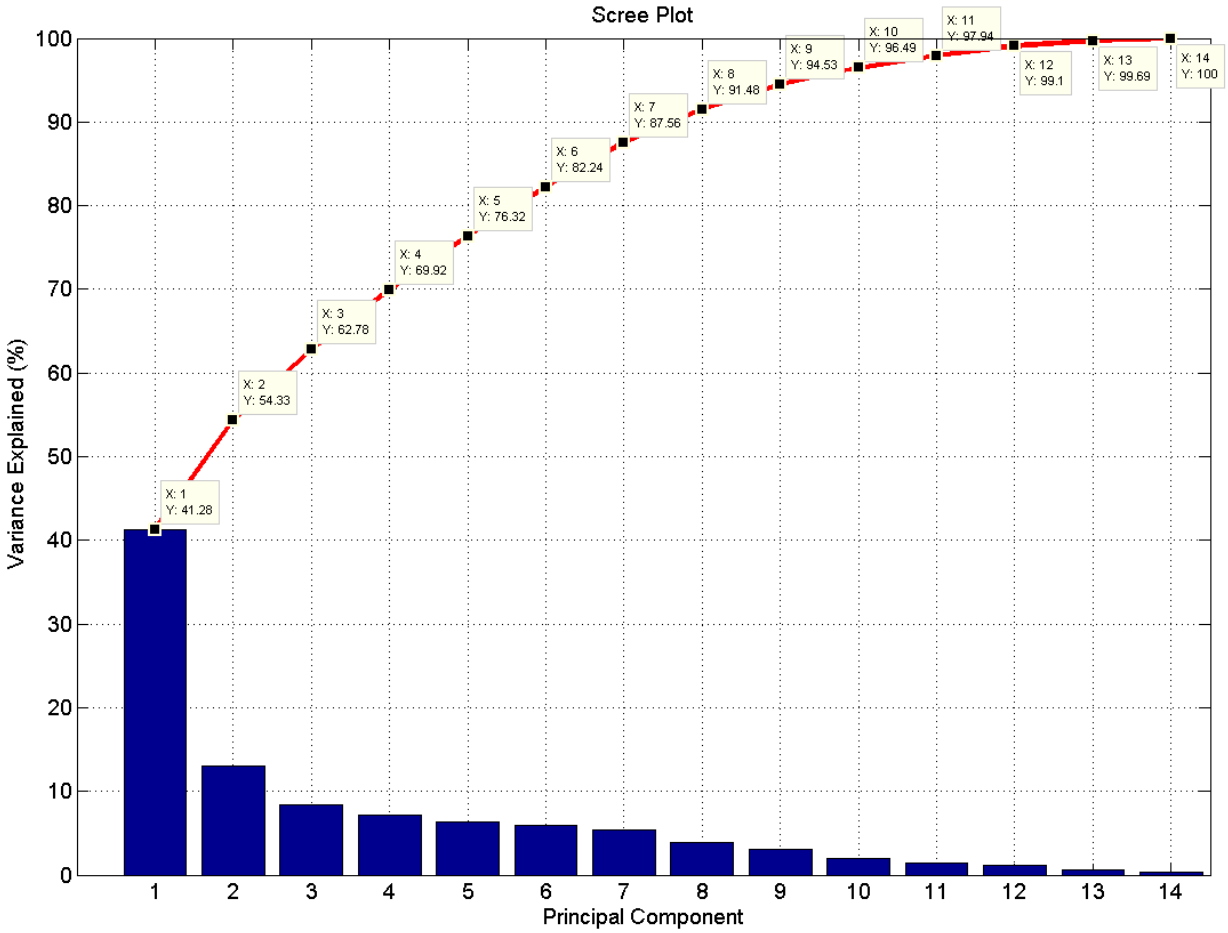



**Figure 4. Scree plot depicting the principal components variance explanation**

Table 2 presents the variance explanation and the sum of variance explanation for all of the principal components.

As seen in the figure and the table the percentage explained is 54.3 with two principal components and 62.8 with three principal components. An explanation of more than 50 % is assumed to be enough to make some quick conclusions from the created biplots in two dimensions, in paragraph 5.4.1 Biplots at page 39

**Table 2. Principal component variance explanation and sum of variance explanation**

| Principal Component | Variance explanation | Sum of variance explanation |
|---|---|---|
| 1 | 41.3 | 41.3 |
| 2 | 13.1 | 54.3 |
| 3 | 8.4 | 62.8 |
| 4 | 7.1 | 69.9 |
| 5 | 6.4 | 76.3 |
| 6 | 5.9 | 82.2 |
| 7 | 5.3 | 87.6 |
| 8 | 3.9 | 91.5 |
| 9 | 3.1 | 94.5 |
| 10 | 2.0 | 96.5 |
| 11 | 1.5 | 97.9 |
| 12 | 1.2 | 99.1 |
| 13 | 0.6 | 99.7 |
| 14 | 0.3 | 100.0 |

## 5.3 Test of stability

The Generated Index is created with fewer principal components in order to study the stability of the model. Four different systems are studied including all fourteen, nine, five and three principal components, Table 3.

As seen in the Table 3 even with as few as 3 principal components the model is stable and generates roughly the same positions of the countries. However there are of course differences and by using 5 principal components a variation of 76,3 % is obtained which should provide a good enough explanation if not all the 14 components is used which of course is preferable.

**Table 3. Generated Index with 14, 9, 5 and 3 principal components**

| Number of PCs | 14 | 9 | 5 | 3 | 14 | 9 | 5 | 3 |
|---|---|---|---|---|---|---|---|---|
| Variance Explained | 100 % | 94.5 % | 76.3 % | 62.8 % | | | | |
| Country | | | | | | | | |
| Iceland | 100.0 | 100.0 | 100.0 | 100.0 | 1 | 1 | 1 | 1 |
| Norway | 79.2 | 78.6 | 76.9 | 77.6 | 2 | 2 | 2 | 2 |
| Luxembourg | 70.0 | 70.0 | 69.3 | 69.4 | 3 | 3 | 3 | 4 |
| Canada | 69.7 | 69.5 | 68.6 | 69.3 | 4 | 4 | 5 | 5 |
| Finland | 69.3 | 69.1 | 68.6 | 70.0 | 5 | 5 | 4 | 3 |
| United States | 68.5 | 68.4 | 67.3 | 68.3 | 6 | 6 | 6 | 6 |
| United Arab Emirates | 67.4 | 67.9 | 66.1 | 64.9 | 7 | 7 | 7 | 9 |
| Sweden | 66.8 | 66.5 | 65.8 | 67.6 | 8 | 8 | 8 | 7 |
| Singapore | 64.4 | 64.4 | 64.0 | 65.0 | 9 | 9 | 9 | 8 |
| Denmark | 64.0 | 64.0 | 61.6 | 63.0 | 10 | 10 | 11 | 12 |
| Australia | 63.6 | 64.0 | 62.8 | 63.6 | 11 | 11 | 10 | 10 |
| Netherlands | 63.3 | 63.3 | 61.6 | 63.4 | 12 | 12 | 12 | 11 |
| New Zealand | 61.5 | 61.8 | 59.8 | 60.8 | 13 | 13 | 13 | 14 |
| Switzerland | 60.9 | 61.0 | 58.7 | 60.5 | 14 | 15 | 15 | 15 |
| United Kingdom | 60.7 | 61.0 | 59.0 | 60.4 | 15 | 14 | 14 | 16 |
| Japan | 59.3 | 59.2 | 57.7 | 61.0 | 16 | 17 | 16 | 13 |
| Germany | 59.2 | 59.3 | 57.3 | 59.7 | 17 | 16 | 18 | 17 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Korea, Rep.** | 58.4 | 58.5 | 57.6 | 59.1 | 18 | 18 | 17 | 18 |
| **Austria** | 57.4 | 57.8 | 55.6 | 57.1 | 19 | 20 | 20 | 19 |
| **Italy** | 57.3 | 57.9 | 54.6 | 56.3 | 20 | 19 | 22 | 20 |
| **Bahrain** | 56.5 | 56.9 | 56.2 | 56.1 | 21 | 21 | 19 | 21 |
| **Belgium** | 55.9 | 56.2 | 54.2 | 56.0 | 22 | 22 | 23 | 22 |
| **Ireland** | 55.4 | 56.0 | 55.0 | 55.2 | 23 | 23 | 21 | 23 |
| **France** | 55.2 | 55.4 | 52.9 | 54.4 | 24 | 24 | 24 | 24 |
| **Czech Republic** | 53.1 | 53.4 | 52.0 | 53.4 | 25 | 25 | 25 | 25 |
| **Israel** | 52.8 | 53.0 | 51.4 | 53.4 | 26 | 26 | 26 | 26 |
| **Portugal** | 49.2 | 49.6 | 47.3 | 50.3 | 27 | 27 | 29 | 28 |
| **Slovenia** | 49.0 | 49.1 | 47.9 | 49.8 | 28 | 28 | 28 | 29 |
| **Malaysia** | 48.1 | 48.4 | 47.3 | 49.3 | 29 | 31 | 30 | 31 |
| **Cyprus** | 48.1 | 48.7 | 46.8 | 47.7 | 30 | 29 | 31 | 33 |
| **Estonia** | 47.8 | 48.6 | 49.8 | 51.1 | 31 | 30 | 27 | 27 |
| **Greece** | 47.7 | 48.2 | 46.7 | 48.3 | 32 | 32 | 32 | 32 |
| **Spain** | 47.5 | 47.5 | 46.2 | 49.5 | 33 | 33 | 33 | 30 |
| **Slovak Republic** | 45.4 | 45.6 | 44.7 | 46.0 | 34 | 34 | 34 | 34 |
| **Croatia** | 45.1 | 45.5 | 43.7 | 45.2 | 35 | 35 | 35 | 37 |
| **Hungary** | 43.7 | 44.1 | 43.3 | 45.9 | 36 | 36 | 36 | 35 |
| **Jamaica** | 42.4 | 42.8 | 38.2 | 39.8 | 37 | 37 | 45 | 44 |
| **Poland** | 42.1 | 42.4 | 41.3 | 42.7 | 38 | 38 | 38 | 38 |
| **Latvia** | 41.4 | 41.7 | 40.6 | 39.8 | 39 | 40 | 39 | 43 |
| **Lithuania** | 41.3 | 41.8 | 41.9 | 42.4 | 40 | 39 | 37 | 39 |
| **Russian Federation** | 41.1 | 41.5 | 39.0 | 39.2 | 41 | 41 | 43 | 46 |
| **Brunei Darussalam** | 40.9 | 41.0 | 39.9 | 45.9 | 42 | 43 | 40 | 36 |
| **Bulgaria** | 40.8 | 41.1 | 39.7 | 40.3 | 43 | 42 | 41 | 41 |
| **Chile** | 39.8 | 40.4 | 39.0 | 41.4 | 44 | 44 | 44 | 40 |
| **Macedonia, FYR** | 39.2 | 39.5 | 37.6 | 39.5 | 45 | 46 | 47 | 45 |
| **Saudi Arabia** | 39.2 | 39.7 | 37.9 | 40.3 | 46 | 45 | 46 | 42 |
| **Trinidad and To-** | 38.1 | 38.9 | 39.1 | 38.7 | 47 | 47 | 42 | 47 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **bago** | | | | | | | | |
| **Thailand** | 38.0 | 38.1 | 36.3 | 38.4 | 48 | 49 | 48 | 48 |
| **Uruguay** | 37.9 | 38.1 | 35.1 | 37.7 | 49 | 48 | 51 | 49 |
| **Ukraine** | 37.7 | 37.9 | 35.4 | 35.5 | 50 | 50 | 50 | 52 |
| **Oman** | 36.4 | 37.0 | 35.5 | 37.5 | 51 | 51 | 49 | 50 |
| **Romania** | 36.3 | 36.9 | 32.8 | 33.8 | 52 | 52 | 56 | 56 |
| **Kazakhstan** | 35.5 | 35.7 | 34.4 | 33.0 | 53 | 53 | 52 | 60 |
| **Argentina** | 35.1 | 35.6 | 33.4 | 36.3 | 54 | 54 | 54 | 51 |
| **Jordan** | 34.9 | 35.0 | 32.9 | 33.2 | 55 | 56 | 55 | 57 |
| **Turkey** | 34.3 | 35.1 | 28.9 | 29.7 | 56 | 55 | 68 | 70 |
| **Tunisia** | 34.1 | 34.5 | 32.7 | 33.0 | 57 | 58 | 57 | 59 |
| **Panama** | 34.0 | 34.6 | 33.7 | 34.3 | 58 | 57 | 53 | 54 |
| **Mexico** | 33.8 | 34.1 | 32.0 | 35.1 | 59 | 59 | 58 | 53 |
| **Moldova** | 32.8 | 33.1 | 30.0 | 30.8 | 60 | 60 | 63 | 65 |
| **Lebanon** | 32.2 | 32.7 | 30.1 | 30.9 | 61 | 61 | 61 | 64 |
| **Iran, Islamic Rep.** | 31.8 | 31.9 | 28.5 | 29.0 | 62 | 63 | 70 | 73 |
| **Kyrgyz Republic** | 31.7 | 31.6 | 28.4 | 28.7 | 63 | 65 | 71 | 74 |
| **Morocco** | 31.7 | 32.0 | 30.4 | 31.0 | 64 | 62 | 60 | 63 |
| **Belarus** | 31.2 | 31.7 | 23.3 | 23.9 | 65 | 64 | 86 | 91 |
| **Syrian Arab Republic** | 31.2 | 31.1 | 27.7 | 29.6 | 66 | 67 | 73 | 71 |
| **Bosnia and Herzegovina** | 30.7 | 31.2 | 30.9 | 32.2 | 67 | 66 | 59 | 61 |
| **Algeria** | 30.4 | 30.8 | 27.5 | 28.3 | 68 | 69 | 74 | 78 |
| **Georgia** | 30.2 | 30.8 | 30.1 | 30.0 | 69 | 68 | 62 | 67 |
| **Egypt, Arab Rep.** | 30.0 | 30.0 | 27.4 | 27.7 | 70 | 72 | 75 | 80 |
| **Colombia** | 30.0 | 30.7 | 28.4 | 31.3 | 71 | 70 | 72 | 62 |
| **Costa Rica** | 29.8 | 30.2 | 29.8 | 34.3 | 72 | 71 | 65 | 55 |
| **South Africa** | 29.0 | 29.5 | 27.4 | 27.8 | 73 | 74 | 76 | 79 |
| **Albania** | 28.9 | 29.6 | 29.4 | 30.4 | 74 | 73 | 66 | 66 |
| **Armenia** | 28.4 | 28.7 | 29.1 | 26.9 | 75 | 76 | 67 | 83 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **China** | 28.3 | 28.8 | 29.9 | 29.4 | 76 | 75 | 64 | 72 |
| **Philippines** | 28.2 | 28.7 | 28.6 | 33.1 | 77 | 77 | 69 | 58 |
| **Paraguay** | 27.6 | 28.0 | 24.8 | 28.4 | 78 | 78 | 82 | 76 |
| **Sri Lanka** | 27.2 | 27.3 | 25.1 | 26.6 | 79 | 80 | 80 | 85 |
| **Guatemala** | 26.9 | 27.3 | 25.1 | 26.9 | 80 | 79 | 79 | 82 |
| **El Salvador** | 26.8 | 27.2 | 26.1 | 29.8 | 81 | 81 | 77 | 69 |
| **Dominican Republic** | 26.4 | 26.7 | 24.9 | 26.8 | 82 | 82 | 81 | 84 |
| **Ecuador** | 25.8 | 26.3 | 25.4 | 29.8 | 83 | 83 | 78 | 68 |
| **Mongolia** | 25.3 | 25.8 | 24.0 | 24.0 | 84 | 85 | 84 | 90 |
| **Gabon** | 25.2 | 26.2 | 22.8 | 25.4 | 85 | 84 | 88 | 87 |
| **Nicaragua** | 24.9 | 25.4 | 23.4 | 25.9 | 86 | 86 | 85 | 86 |
| **Honduras** | 24.3 | 24.7 | 22.5 | 25.0 | 87 | 87 | 89 | 88 |
| **Peru** | 24.1 | 24.5 | 24.4 | 28.5 | 88 | 88 | 83 | 75 |
| **Venezuela, RB** | 23.2 | 23.8 | 20.9 | 27.2 | 89 | 89 | 91 | 81 |
| **Vietnam** | 22.8 | 23.3 | 23.0 | 24.8 | 90 | 90 | 87 | 89 |
| **Bolivia** | 21.5 | 22.1 | 20.4 | 23.0 | 91 | 91 | 92 | 92 |
| **Pakistan** | 21.4 | 21.9 | 18.4 | 17.3 | 92 | 92 | 94 | 96 |
| **Brazil** | 19.9 | 20.1 | 20.4 | 28.4 | 93 | 93 | 93 | 77 |
| **Nepal** | 19.5 | 20.0 | 17.0 | 17.4 | 94 | 94 | 95 | 95 |
| **Azerbaijan** | 18.8 | 19.0 | 22.3 | 20.0 | 95 | 95 | 90 | 94 |
| **Indonesia** | 17.7 | 18.1 | 15.4 | 20.8 | 96 | 96 | 96 | 93 |
| **Cote d'Ivoire** | 16.7 | 16.9 | 13.7 | 16.1 | 97 | 97 | 98 | 97 |
| **Togo** | 16.3 | 16.7 | 13.2 | 15.0 | 98 | 98 | 101 | 100 |
| **India** | 15.3 | 16.1 | 15.0 | 14.5 | 99 | 99 | 97 | 101 |
| **Benin** | 14.3 | 14.8 | 12.0 | 11.7 | 100 | 100 | 103 | 105 |
| **Yemen, Rep.** | 14.1 | 14.6 | 11.3 | 12.8 | 101 | 101 | 105 | 102 |
| **Bangladesh** | 13.9 | 14.4 | 13.6 | 15.1 | 102 | 102 | 99 | 99 |
| **Senegal** | 13.8 | 14.2 | 12.1 | 12.1 | 103 | 103 | 102 | 104 |
| **Botswana** | 13.8 | 13.9 | 13.2 | 15.8 | 104 | 105 | 100 | 98 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Sudan** | 13.4 | 14.0 | 11.7 | 10.2 | 105 | 104 | 104 | 109 |
| **Kenya** | 13.3 | 13.5 | 10.9 | 11.6 | 106 | 106 | 107 | 106 |
| **Tanzania** | 13.0 | 13.1 | 11.2 | 10.4 | 107 | 107 | 106 | 108 |
| **Ghana** | 12.5 | 12.8 | 9.6 | 10.8 | 108 | 108 | 109 | 107 |
| **Zambia** | 12.1 | 12.0 | 7.8 | 6.5 | 109 | 109 | 112 | 112 |
| **Namibia** | 11.3 | 11.9 | 10.2 | 12.3 | 110 | 110 | 108 | 103 |
| **Cameroon** | 10.9 | 11.4 | 9.1 | 8.5 | 111 | 111 | 110 | 110 |
| **Ethiopia** | 10.2 | 10.9 | 8.7 | 6.4 | 112 | 112 | 111 | 113 |
| **Nigeria** | 7.6 | 8.2 | 4.1 | 1.7 | 113 | 113 | 114 | 114 |
| **Cambodia** | 5.7 | 6.3 | 7.1 | 7.3 | 114 | 114 | 113 | 111 |
| **Mozambique** | 0.0 | 0.0 | 0.0 | 0.0 | 115 | 115 | 115 | 115 |

## 5.4 Sensitivity analysis

### 5.4.1 Biplots

Several interesting points can be seen in the biplot. Starting of with the variables. As described before if the variables are close to each other they have much in common and if they are way apart they differ allot. The variable Mortality rate is in the opposite direction as Life expectancy while Population growth is close to the Mortality rate which means that countries with a high Population growth also have high Mortality rate. Immunization to measles is fairly close to Life expectancy which points out that they are correlated and indicates that with a high immunization to measles the possibility for a long life increases. GDP growth and Inflation are almost overlapping showing a strong correlation between these two economic variables. Variables like High technology exports, Electric power consumption, Internet users, Mobile cellular subscriptions CO2 emissions are closely connected with a fairly close connection to Renewable internal freshwater resources. All of the later factors are closely connected to a developed country thus it is reasonable to assume that developed countries places closely to these variables while the developing countries should place closer to the variables Mortality rate and Population Growth.

Time required to start a business are fairly close connected with GDP growth and Inflation but not with the previously larger cluster of variables. The paved Roads is somewhat of a wild card closest connected to Immunization to measles and conclusions hereby are not realized easily. It is important to keep in mind that the biplot only explains the 54.3 % of the variation but still it gives a sort of indication of how the variables are connected or not.

To clarify the biplot the following examples can be studied. As seen in the biplots (Figure 5) on page 41 the mortality rate factor is a long vector with several countries within a small angle from the vector and thus have a large correlation.

It's shown in the plot (Figure 5) that countries like Cameroon, Ethiopia, Nigeria, Cambodia and Mozambique are grouped together and have strong correlations to the Mortality rate, under-5 (per 1,000). They have all values of 100 or above which only 16 countries in total have (values =>95 is counted as 100 or above). Table 4 below shows countries with Mortality rate, under-5 (per 1,000), of $\geq 95$.

**Table 4. Countries with Mortality rate, under-5 (per 1,000), of >=95**

| Country | Mortality rate, under-5 (per 1,000) |
|---|---|
| **Nigeria** | 195.3 |
| **Mozambique** | 174.6 |
| **Zambia** | 173.8 |
| **Cameroon** | 149.2 |
| **Benin** | 132.2 |
| **Ethiopia** | 132.1 |
| **Cote d'Ivoire** | 129.9 |
| **Tanzania** | 127.6 |
| **Senegal** | 120.3 |
| **Kenya** | 119.7 |
| **Ghana** | 113.9 |
| **Sudan** | 110.9 |
| **Togo** | 107.8 |
| **Pakistan** | 97.0 |
| **Cambodia** | 96.4 |

In the right part of the plot Iceland is the outlier of the system reaching far beyond any other country. The other closely grouped countries to the right are Norway, Luxembourg, Cananda Finland, United States, Sweden, United Arab Emirates, Switzerland, Ireland, Singapore and Austria. They are gathered around the Internet users, Mobile cellular subscriptions, Electric power consumption, High-technology exports, and CO2 emissions.

Besides of explaining the variables connection to each other the biplot also gives indication of how, in this case, the countries are correlated to factors. As explained above the developed countries clustered together as well as the developing countries.

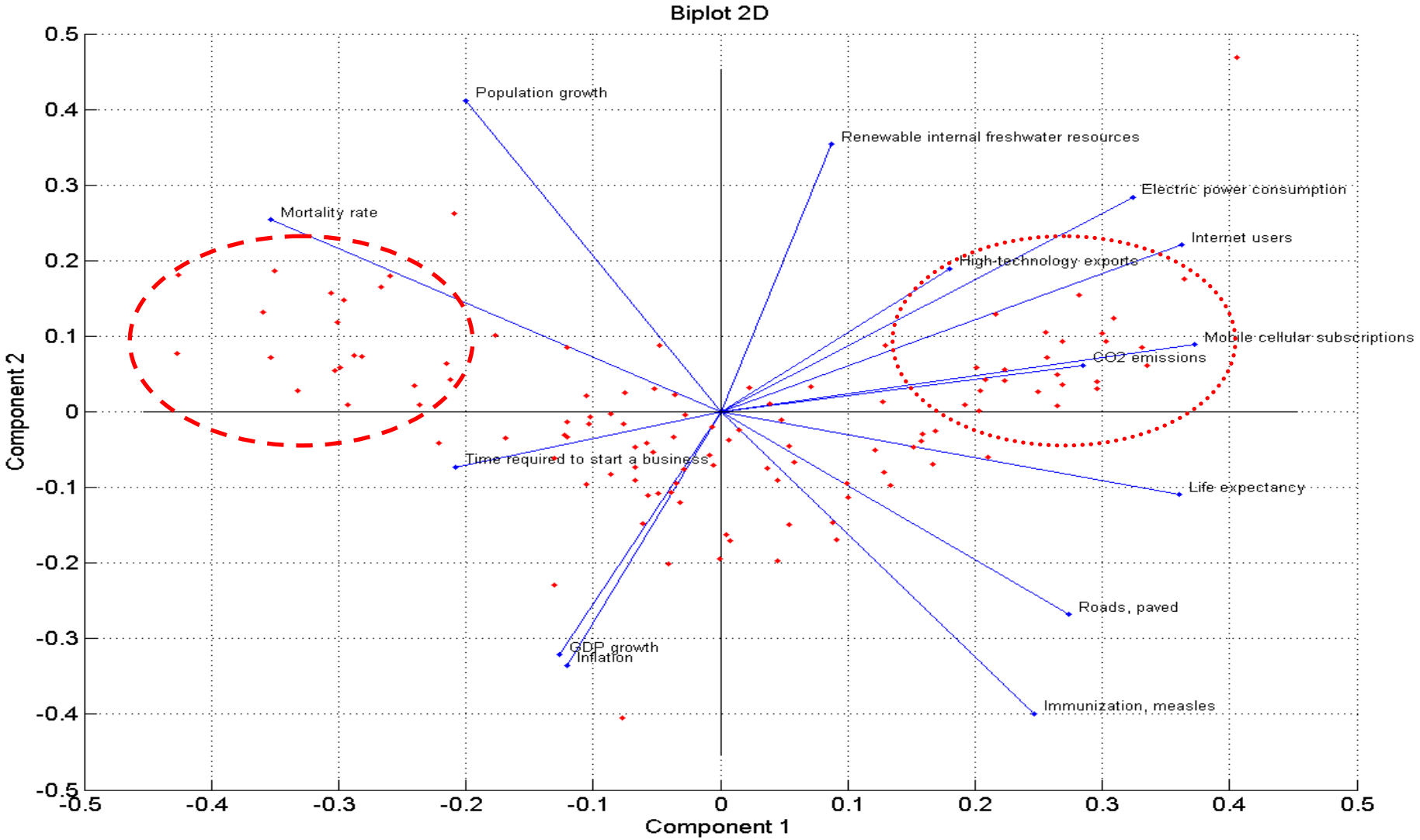



**Figure 5. Biplot in two dimensions**

### 5.4.2 Dendrogram

The dendrogram created with average suclidean distance confirms that Iceland is different from the others and only groups Iceland together with itself. The theory of dendrograms is explained in the Appendix at page 61.

Three clearly visible groups emerge from the dendrogram and these are shown in Table 5 below. Close up images for each cluster can be found in the Appendix at page 74-77.

The dendrogram clearly points out three clusters of countries similar to each other. The dendrogram also clarifies that Iceland is a country unlike all others as it is placed in the far end at an own branch. The first cluster (Cluster 1) is made up mainly by African countries with some Asian countries as India and Bangladesh. The second (Cluster 2) cluster consists of European countries, North America, Australia and Japan. The third cluster (Cluster 3) consist of South America and Asia some African and European countries.

As expected developed countries in Europe, North Africa and the country Australia clusters together and developing countries in Africa and Asia clusters together. Looking at the GI ranking Cluster 2 groups from 2-40 (with three countries in the region 43-49), Cluster 3 groups from 41-90, and Cluster 1 groups from 91-115 (with two countries in the region 73-85). In conclusion the dendrogram suggest that there are in general three categories 1-40, 41-90 and 91-115.

**Table 5. Clusters from dendrogram**

| Cluster 1 | GI rank | Cluster 2 | GI rank | Cluster 3 | GI rank |
|---|---|---|---|---|---|
| South Africa | 73 | Norway | 2 | Russian Federation | 41 |
| Gabon | 85 | Canada | 4 | Chile | 44 |
| Bolivia | 91 | Finland | 5 | Saudi Arabia | 46 |
| Pakistan | 92 | United States | 6 | Trinidad and Tobago | 47 |
| Nepal | 94 | Sweden | 8 | Thailand | 48 |
| Cote d Ivorie | 97 | Singapore | 9 | Ukraine | 50 |
| Togo | 98 | Denmark | 10 | Oman | 51 |
| India | 99 | Australia | 11 | Romania | 52 |
| Benin | 100 | Netherlands | 12 | Kazakhstan | 53 |

| | | | | | |
|---|---|---|---|---|---|
| Yemen, Rep. | 101 | New Zealand | 13 | Argentina | 54 |
| Bangladesh | 102 | Switzerland | 14 | Jordan | 55 |
| Senegal | 103 | United Kingdom | 15 | Turkey | 56 |
| Botswana | 104 | Japan | 16 | Tunisia | 57 |
| Sudan | 105 | Germany | 17 | Panama | 58 |
| Kenya | 106 | Korea, Rep. | 18 | Mexico | 59 |
| Tanzania | 107 | Austria | 19 | Moldova | 60 |
| Ghana | 108 | Italy | 20 | Lebanon | 61 |
| Zambia | 109 | Belgium | 22 | Iran, Islamic Rep. | 62 |
| Namibia | 110 | Ireland | 23 | Kyrgyz Republic | 63 |
| Cameroon | 111 | France | 24 | Morocco | 64 |
| Ethiopia | 112 | Czech Republic | 25 | Syrian Arab Republic | 66 |
| Nigeria | 113 | Israel | 26 | Bosnia and Herzegovina | 67 |
| Cambodia | 114 | Portugal | 27 | Algeria | 68 |
| Mozambique | 115 | Slovenia | 28 | Georgia | 69 |
| | | Malaysia | 29 | Egypt, Arab Rep. | 70 |
| | | Cyprus | 30 | Colombia | 71 |
| | | Estonia | 31 | Albania | 74 |
| | | Greece | 32 | Armenia | 75 |
| | | Spain | 33 | China | 76 |
| | | Slovak Republic | 34 | Paraguay | 78 |
| | | Croatia | 35 | Sri Lanka | 79 |
| | | Hungary | 36 | Guatemala | 80 |
| | | Jamaica | 37 | El Salvador | 81 |
| | | Poland | 38 | Dominican Republic | 82 |
| | | Latvia | 39 | Ecuador | 83 |
| | | Lithuania | 40 | Mongolia | 84 |
| | | Bulgaria | 43 | Nicaragua | 86 |
| | | Macedonia, FYr | 45 | Honduras | 87 |
| | | Uruguay | 49 | Peru | 88 |
| | | | | Vietnam | 90 |

### 5.4.3 Studies of regressions and residuals

#### *5.4.3.1 Linear regression in 2D*

A linear regression with two indexes has been created for three cases in order to find linearity:

1. HDI vs. Generated index
2. DI vs. Generated index
3. GNI per capita (PPP) vs. Generated index

Figure 6 shows a close connection between the GI and HDI which is obvious when studying the variables which makes up the HDI. Life expectancy is also included in GI along with some sort of BNP per capita, with the BNP growth. However the "knowledge"-factor is not included in the GI. The 95% confidence interval is much narrower than the one with the DI. In order to see more exactly how well the linear regression fits the data the residuals are studied. The linear correlation between the GNI per capita (PPP) and Generated Index is less than for the one with HDI but higher than for the one with DI.

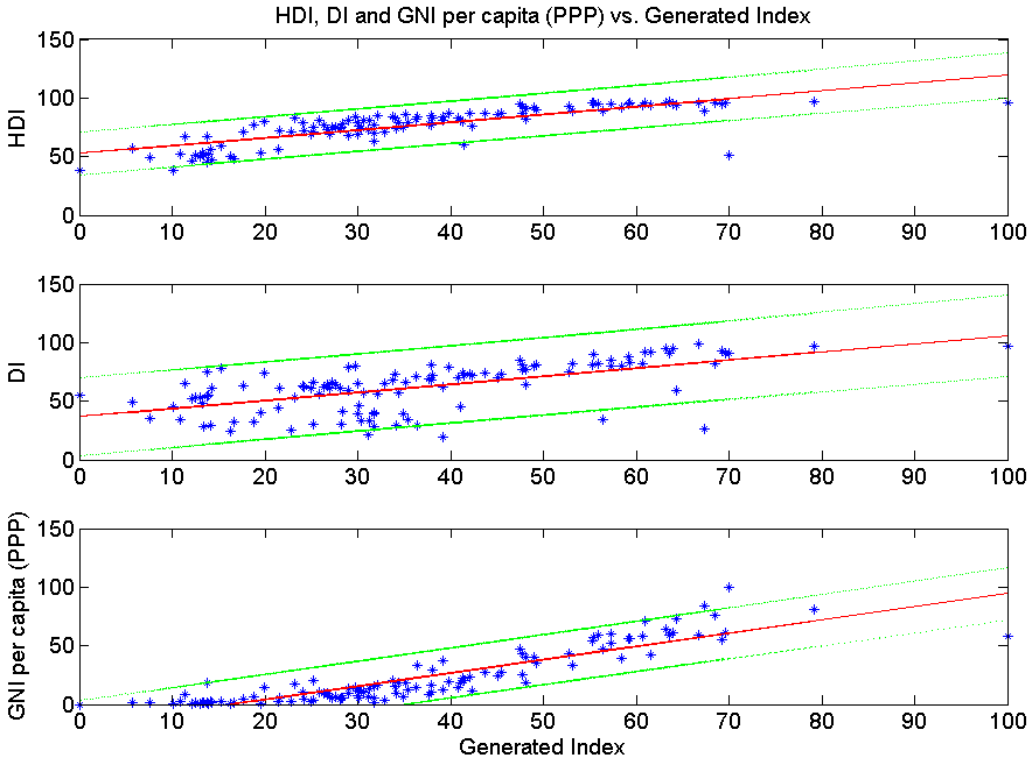



**Figure 6. Linear regressions with HDI vs. GI, DI vs. GI and GNI per capita (PPP) vs. GI.**

### *5.4.3.2 Residual Analysis*

The residuals to the linear regression have been calculated and plotted in the graphs below, Figure 7 . As seen in figure the residuals meet the assumptions of a linear correlation. (Olbjer 2000). The residuals are also shown to be normally distributed and the linear regression is a good fit.

This indicates that there is a correlation between the GI and HDI, between the GI and DI and between the GI and GDI.

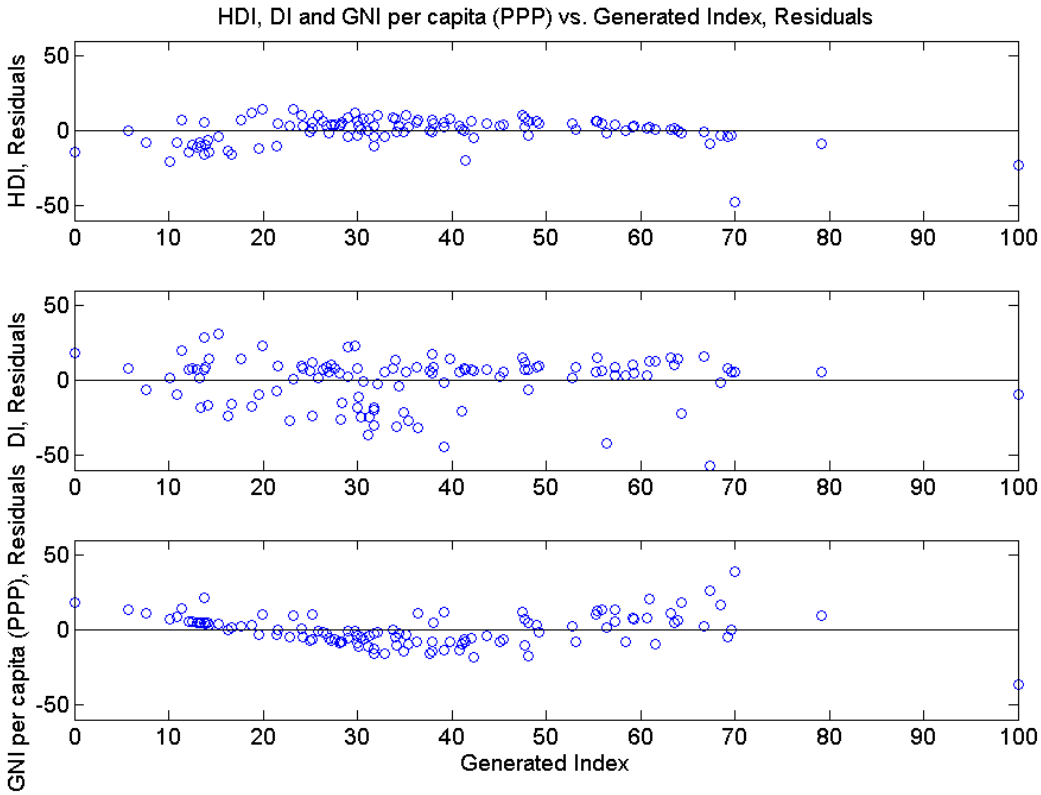



**Figure 7. Residuals for the three linear regression models: HDI vs GI, DI vs. GI and GNI per capita (PPP) vs. GI.**

To study the distribution of the residuals they have been plotted in a histogram with a superimposed fitted normal density, Figure 8. The shapes of the bell curves are pointiest at the "HDI vs. GI" followed by "GNI per capita (PPP) vs. GI" and last the "DI vs. GI":

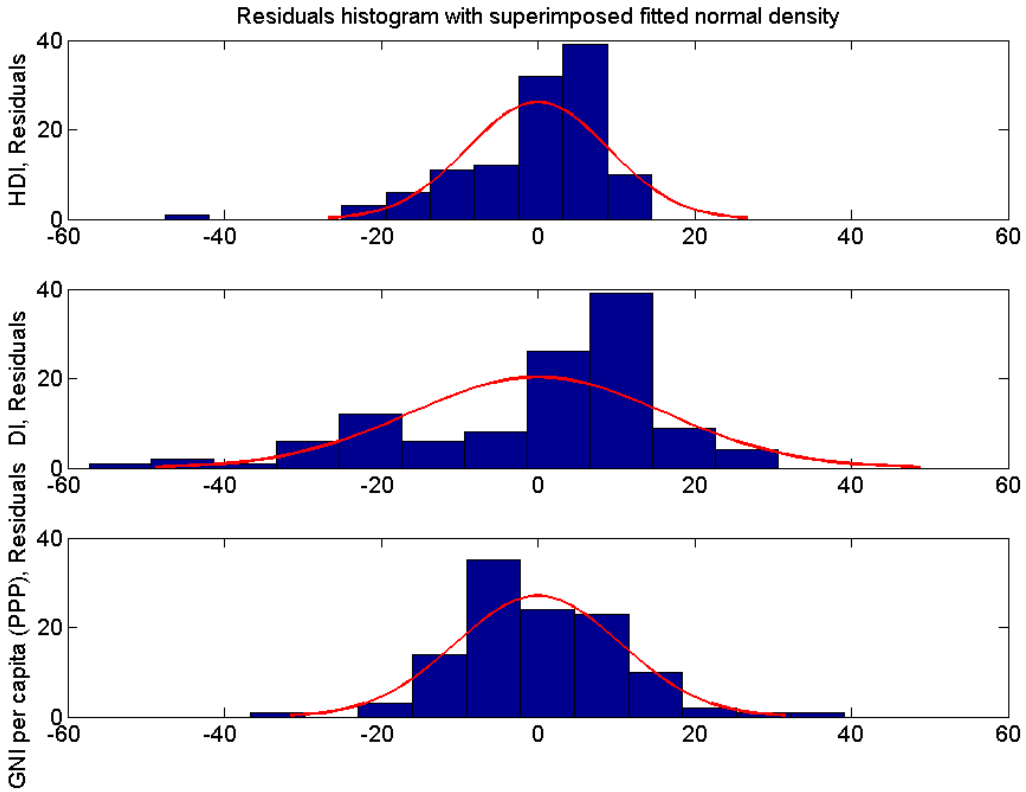



**Figure 8. Residuals histogram with superimposed fitted normal density**

### 5.4.4 Linear regression in 3D

A linear regression has been created in the three dimensional plot in order to find a linear correlation. The strength of the three dimensional plot is that it can be deducted if there is a correlation between three variables and not merely between two. The distance from the regression is also plotted in a brighter blue to clearly point out if some countries differs heavily from the regression.

Figure 9 shows a linear connection between GI, HDI and DI. The most extreme points are from Iceland and Luxembourg which both are small countries with fairly small economies. It is clear that there is a connection between the three indexes and this will strengthen the possibility to exchange the DI and HDI for the GI in the present.

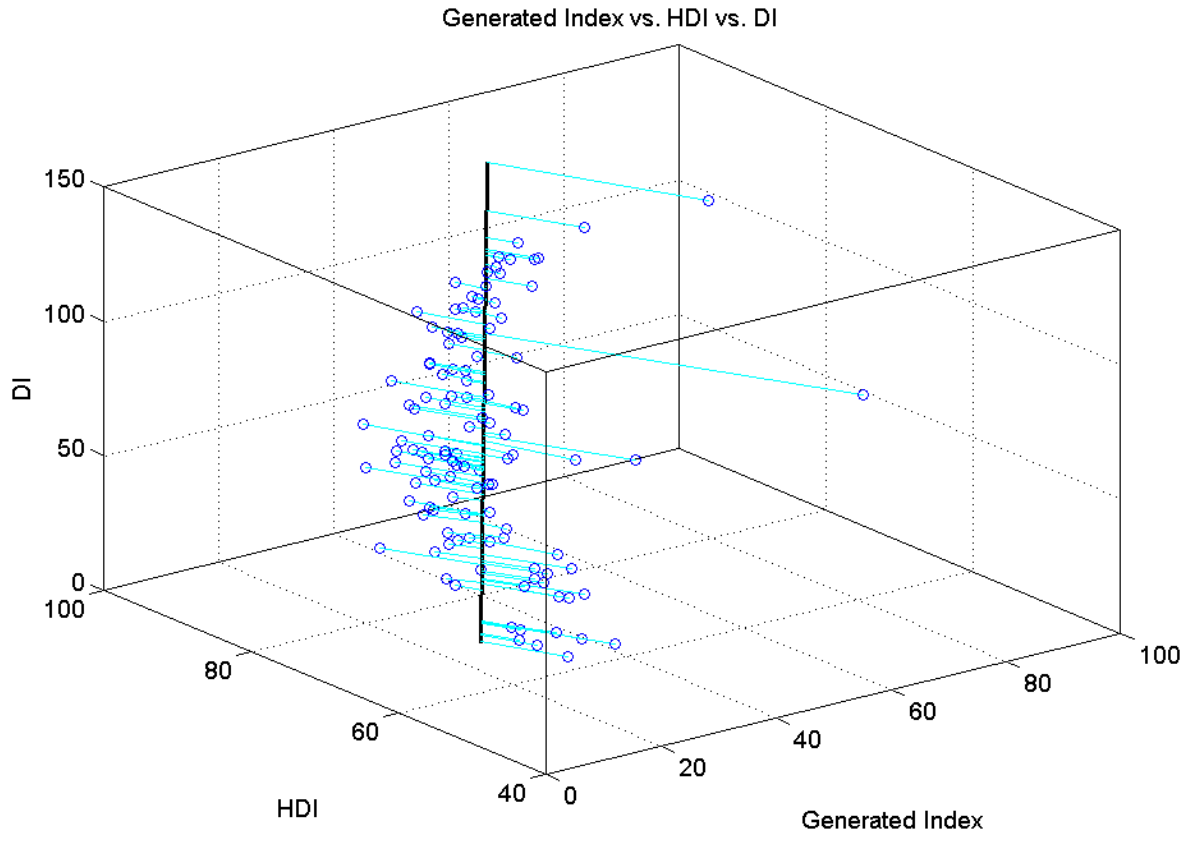



**Figure 9. 3D linear regression with DI, HDI and Generated Index**

### 5.4.4.1 *Residual analysis*

The residuals to the linear regression have been calculated and plotted in the graphs below, Figure 10 and Figure 11. As seen in figure the residuals meet the assumptions of a linear correlation. (Olbjer 2000):

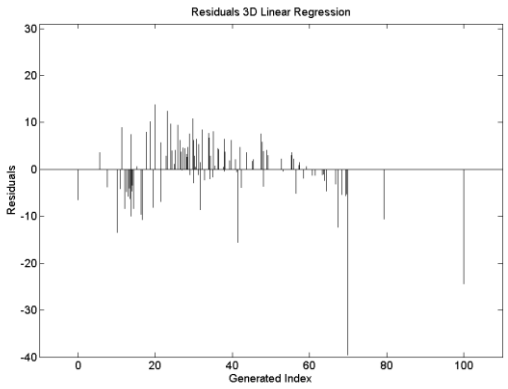



**Figure 10 - Residuals for the 3D linear regression, bar plot**

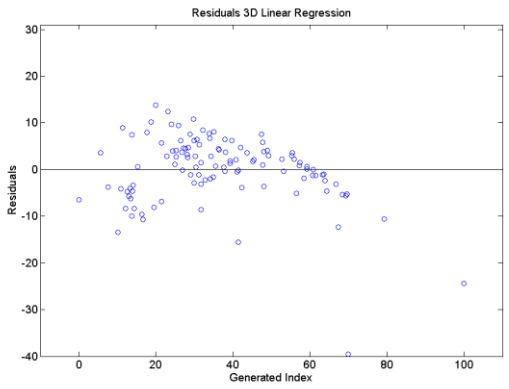



**Figure 11 - Residuals for the 3D linear regression**

Figure 12 shows that the residuals are normally distributed.

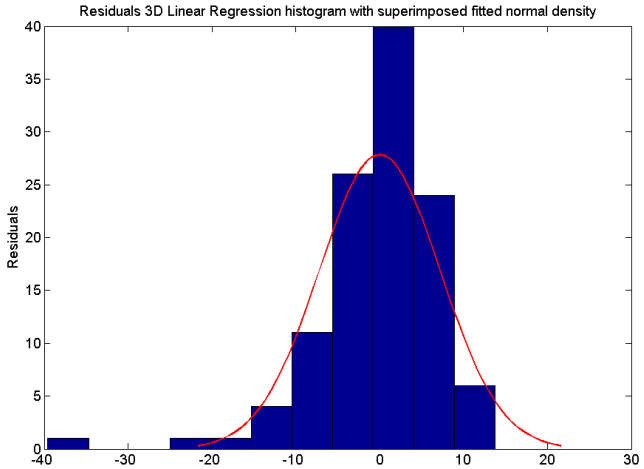



**Figure 12. Residuals for the 3D Linear Regression histogram with superimposed fitted normal density.**

## 5.4.5 Engle's Autoregressive conditional heteroscedasticity (ARCH) model

Arch test for all residuals are conducted in MATLAB. In all residual plots the function returns a value of zero which indicates acceptance of the null hypothesis that no ARCH effect exist (i.e., homoscedasticity). This assures that there is no heteroscedasticity in the regressions.

## 5.5 Time series analysis

In order to study how the Generated Index varies over time (1960-2008) a study of 11 different countries is concluded. The time series includes the unscaled GI so that each and every country only is compared to itself. This should tell how the welfare in the country changes over the years and it will be possible to see if the welfare have increased or declined during the period of study. With the previously calculated loadings (se paragraph A2 Principal Component Analysis: finding the components at page 59 in the Appendix) the same weights are used when the unscaled GI is calculated in the time series analysis. Countries from both the bottom, middle and the top of the unscaled GI list are chosen and from all continents (not Antarctica, and Australia).

### 5.5.1 Crisis correlation

In order to see if the index has the possibility to depict the change in welfare during the time span a so called "Crisis correlation" is sought after. In other words; is there a drop in welfare during the period of crisis in the troubled countries? From Wikipedia (Wikipedia 2010) the following dates from 1960-2008 are found to involve crisis in some form, Table 6:

**Table 6. Economic crises from 1960-2008**

| Year | Type of Crisis | Affected Country/Countries |
|---|---|---|
| **1973-1974** | Oil and stock market crisis | Global |
| **1973-1975** | Secondary banking crisis | United Kingdom |
| **1980s** | Debt Crisis beginning in Mexico 1982, *Mexican Weekend* | Latin America |
| **1983** | Bank stock crisis | Israel |
| **1987** | Stock market, *Black Monday* | Global |
| **1989-91** | Savings & Loan crisis | United States |
| **1990** | Collapsing of the asset price bubble | Japan |
| **Early 1990s** | Bank crisis | Sweden & Finland |
| **1992-93** | European Exchange Rate Mechanism (ERM) crisis, *Black Wednesday* | Global |
| **1994-95** | Speculative attack and default on debt, | Mexico |

| | | |
|---|---|---|
| | *Mexican peso crisis* | |
| 1997-98 | Devaluations and banking crises | Asia |
| 1998 | Financial crisis, *Ruble* crisis | Russia |
| 2001 | Internet companies, *Dot-com bubble* | Global |
| 2007-10 | Financial crisis | Global |

### 5.5.2 The overall time series

The following countries have been plotted in Figure 13 over the studied time span: Sweden, United States, Kenya, Mexico, China, Japan, Mozambique, United Kingdom, Brazil, India and Russian Federation.

In general the unscaled Generated Index varies strongly over the years but keeping the crises in mind this is not unexpected**.** The general conclusion of the graphs is that there are more extreme values in the middle of the 1980s and forward which make sense because of the collapse of the stock market 1987 followed by the savings and loan crisis in the United States the following years. And this trend continues during the 1990s and 2000s in correlation with an increased amount of crises.

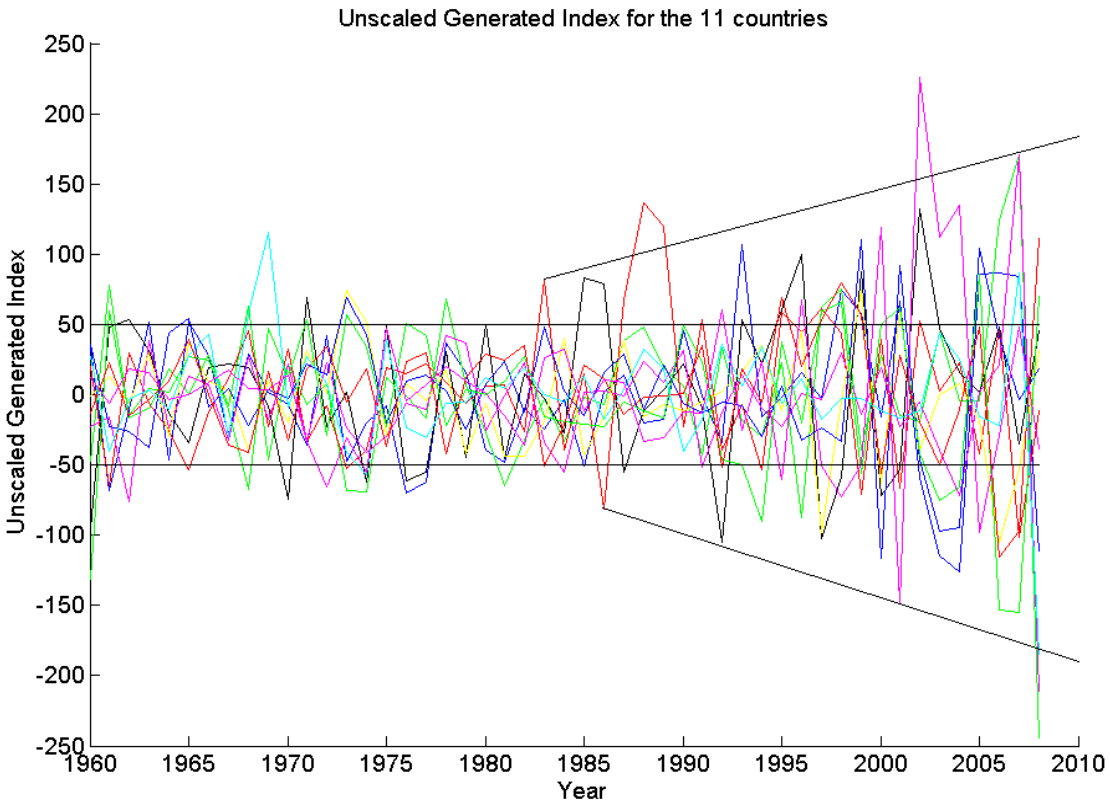



**Figure 13. Unscaled Generated Index for the 11 countries**

### 5.5.3 The specific time series

Besides the total time series the countries have been plotted separately as seen in Table 7. Sweden is included as a reference for the reader in all the figures. The figures can also be found in close up in paragraph A9 Time series plots at page 78.

**Table 7. Countries included in the time series analysis**

| Figure 14 | Figure 15 | Figure 16 | Figure 17 | Figure 18 |
|---|---|---|---|---|
| Sweden | Sweden | Sweden | Sweden | Sweden |
| United States | Kenya | Mexico | China | Japan |
| Mozambique | United Kingdom | Brazil | India | Russian Federation |

In the oil crisis of 1973 Sweden, United States, Mozambique, Japan and Russian Federation drops while United Kingdom is low during the whole period of 1973-1975. Kenya drops first in 1974 and Mexico, Brazil, China and India drops in 1975.

In the crisis of the 1980s Sweden, United States, Mozambique, Kenya, Mexico and Russian Federation drops while Chine, India and the United Kingdom first drops in 1983. Japan is relatively untouched by these crises in the 80s.

The stock market crisis in 1987 affects the United States, Mozambique, Kenya and Russian Federation whilst the United Kingdom, Brazil, Mexico, China, India, Japan and Sweden are relatively untouched.

In the beginning of the 1990s all countries decline except Mexico, Brazil and Russia. In 1996 the United States and Kenya drops followed by Mexico which plummets in 1997. Russian Federation is low during the period of 1997-1999 at the time of the Ruble crisis. China is down in 1999 but not during the devaluations and banking crisis in 1997-1998 in Asia. Russian Federation skyrockets in 2002. The crisis of 2001 is one of the most clearly visible crisis in the graph along with the crisis of 2007 with the most extremes in both directions. Mozambique and Mexico drops in 2001 whilst the United Kingdom, Sweden, Brazil, India and China drops in 2002-2004. The United States is relatively untouched together with Russian Federation and Japan and instead increase more or less constant from 2001 to 2007 and forward with Russian Federation as an exception which drops in 2005. Mozambique, Kenya, Mexico, Brazil, China and India incline in 2008 where as the rest decline.

## Time series of specific countries with Sweden as reference

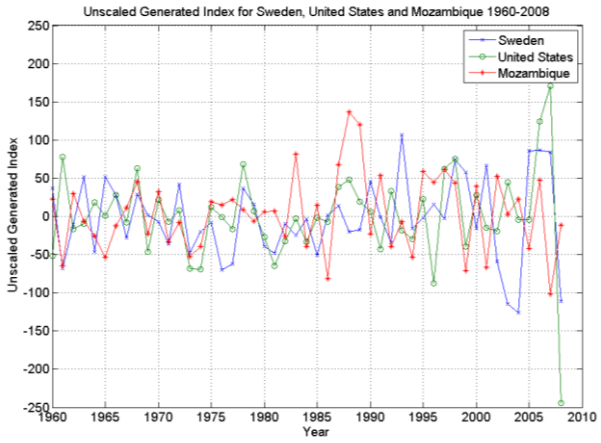



**Figure 14. Unscaled Generated Index for Sweden, United States and Mozambique 1960-2008**

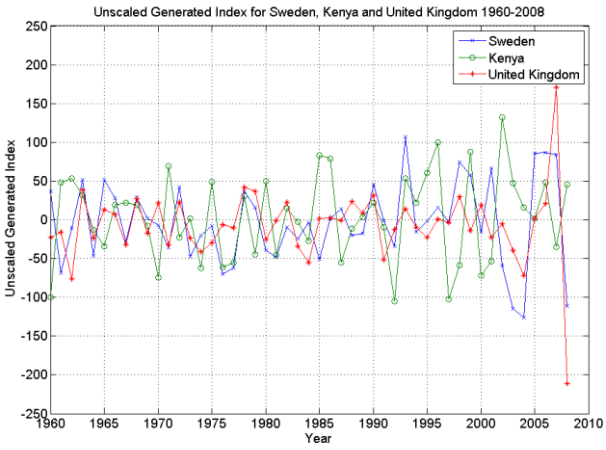



**Figure 15. Unscaled Generated Index for Sweden, Kenya and United Kingdom 1960-2008**

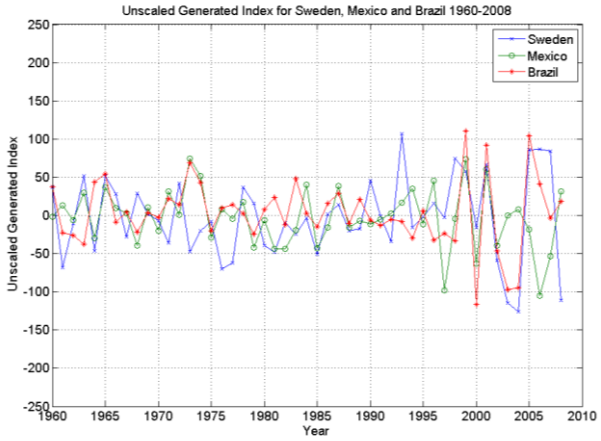



**Figure 16. Unscaled Generated Index for Sweden, Mexico and Brazil 1960-2008**

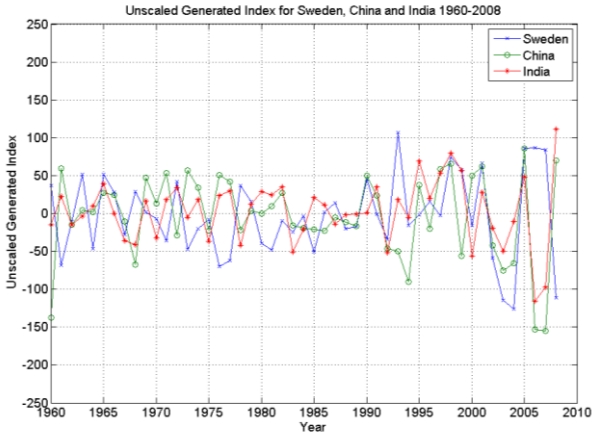



**Figure 17. Unscaled Generated Index for Sweden, China and India 1960-2008**

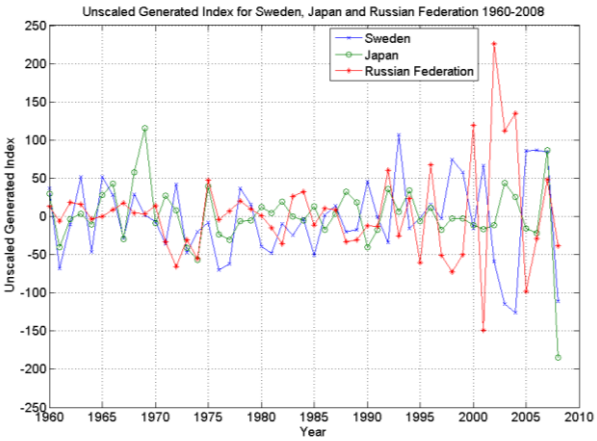



**Figure 18. Unscaled Generated Index for Sweden, Japan and Russian Federation 1960-2008**

A crude linear regression of the countries shows that the welfare of Sweden, United States, Mozambique, Kenya and Russian Federation have risen while Mexico, Brazil, China and Japan have declined. United Kingdom and India are basically unchanged in their Welfare, Table 8. It seems that most countries should have a better welfare today than in the 60s according to HDI for example where all of these countries have a better value today than in the 80s but some countries don't have data for 80s thus making it hard to evaluate.

**Table 8. Results from crude linear regression over the time period 1960-2008**

| Country | Welfare up or down |
|---|---|
| **Sweden** | Up ⇑ |
| **United States** | Up ⇑ |
| **Mozambique** | Up ⇑ |
| **Kenya** | Up ⇑ |
| **United Kingdom** | Unchanged ⇒ |
| **Mexico** | Down ⇓ |
| **Brazil** | Down ⇓ |
| **China** | Down ⇓ |
| **India** | Unchanged ⇒ |
| **Japan** | Down ⇓ |
| **Russian Federation** | Up ⇑ |

# 6 Conclusions

## 6.1 Does the world need the Generated Index?

In the present of 2008 the Generated Index explains the HDI and DI in a fairly good way thus it has the possibility to replace the two of them. However in a longer period of time the index is not to be equal to HDI. As for the DI there is only data for a couple of years and conclusion of this cannot be drawn during a longer period of time. The Generated Index could instead perhaps be used as a crisis predictor of some sort. The correlation with the different crises during the time period 1960-2008 is imminent and with some extended study of the model a better crisis predictor can possibly be created. The benefits of the Generated Index is that it is possible to create it for the period 1960 and forward whereas the HDI only exist from 1980 and for some countries only from as late as 2005. In general this is a problem because of the lack of historic indexes and the world would benefit from a better understanding of our history with more indexes over a longer period of time.

## 6.2 Time series

The time series shows in general two types of countries depending on the time of study, the first is the stable ones that do not have either peaks or bottoms and the other which have both. The stable countries will not have the benefits from the peaks but not the detriments from the drops. As discussed earlier the amount of crises will increase if the fluctuations cannot be controlled and kept within reasonable limits.

## 6.3 The model rather than the index over time

A conclusion of this study is that the model itself, with principal component analysis, has great possibilities to describe the world according to several variables, however it is much dependant on which variables are chosen. If deeper studies of the variables themselves where to be undertaken then the method of choice is believed to work well. Seeing several other fairly new indexes such as the ISEW, GNH and Legatum Prosperity Index which all includes several variables these types of indexes will probably be more common in the near future as the possibility of advanced statistics become more easily conducted. With the use of principal component analysis more advanced index can be generated to better understand the past, current and future welfare of the world.

# 7 References

## 7.1 Literature


1) Abdi, H. and L. J. Williams (2010). Principal Component Analysis. W. I. R. C. Statistics. 2.
2) Adler, N., E. Yazhemsky, et al. (2010). "A framework to measure the relative socio-economic performance of developing countries." Socio-Economic Planning Sciences 44: 73-88.
3) Antony, G. M. and K. V. Rao (2007). "A composite index to explain variations in poverty, health, nutritional status and standard of living: Use of multivariate statistical methods." Public Health 121: 578-587.
4) Barreto, H. and F. M. Howland (2006). Introductory econometrics: using Monte Carlo simulation with Microsoft Excel. New York, Cambridge university press.
5) Basu, R. L. (2005). "Why the Human Development Index Does not Measure up to Ancient Indian Standards." Culture Mandala 6(2): Article 5, no page numbers presented.
6) Beça, P. and R. Santos (2010). "Measuring sustainable welfare: A new approach to the ISEW." Ecological Economics 69: 810–819.
7) Brereton, R. G. (2003). Chemometrics - Data Analysis for the Laboratory and Chemical Plant. West Sussex, England, John Wiley & Sons Ltd.
8) Brodsky, D. A. and D. Rodrik (1981). "Indicators of development and data availability: The case of the PQLI." World Development 9(7): 695-699.
9) Daly, H. E. and J. Cobb (1989). For the common good: redirecting the economy toward community, the environment, and a sustainable future. Boston, Beacon Press.
10) Economist, T. (2008). The Economist Intelligence Unit's Index of Democracy 2008. The Economist. 2010.
11) Engle, R. F. (1983). "Estimates of the Variance of U.S. Inflation Based upon the ARCH Model." Journal of Money, Credit and Banking 15(3): 286-301.
12) Gower, J. C. and D. J. Hand (1996). Biplots. Cornwall, CRC Press.
13) Griswold, D. T. (2004). "Trading Tyranny for Freedom: How Open Markets Till the Soil for Democracy." Trade Policy Analysis 26: 1-16.

14) Kulig, A., H. Kolfoort, et al. (2010). "The case for the hybrid capital approach for the measurement of the welfare and sustainability." Ecological Indicators 10: 118-128.
15) Lawn, P. A. (2003). "A theoretical foundation to support the Index of Sustainable Economic Welfare (ISEW), Genuine Progress Indicator (GPI), and other related indexes." Ecological Economics 44: 105-118.
16) Morris, M. D. (1977). "A physical quality of life index." Urban Ecology 3: 225-240.
17) Neumayer, E. (1999). "THE ISEW – NOT AN INDEX OF SUSTAINABLE ECONOMIC WELFARE." Social Indicators Research 48(1): 77-101.
18) NewEconomicsFoundation (2006). The happy planet index, New Economics Foundation (NEF) 1.
19) Olbjer, L. (2000). Experimentell och industriell statistik. Lund, Matematikcentrum, Lunds Tekniska Högskola.
20) Paterson, M. (2006). Consumption and everyday life. Bristol, Routledge.
21) Perloff, J. M. (2009). Microeconomics. Boston, Pearson
22) Sagara, A. D. and A. Najam (1998). "The human development index: a critical review." Ecological Economics, 25(3): 249-264.
23) Sen, A. K. (1987). On Ethics and Economics. Oxford, Blackwell.
24) Steel, P., J. Schmidt, et al. (2008). "Refining the relationship between personality and subjective well-being." Psychological Bulletin 134: 138–161.
25) Varian, H. R. (2006). Intermediate microeconomics New York and London, Norton.
26) Winton, B. (2009). "Gross national happiness." Asian-Pacific Economic Literature 23(2): 1-16.

## 7.2 Electronic


1) Caplan, B. (2009). "Against the Human Development Index." Retrieved 23 May, 2010, from http://econlog.econlib.org/archives/2009/05/against_the_hum.html#.
2) Dictionary.com. (2010). Retrieved 7 April, 2010, from http://dictionary.reference.com/browse/welfare.
3) LegatumInstitute (2009). Legatum prosperity index. United Kingdom, Legatum Institute. http://www.li.com/
4) Merriam-Webster. (2010). Retrieved 7 April, 2010, from http://www.merriam-webster.com/dictionary/welfare.
5) TheWorldBank. (2010). "WDI, GDF & ADI Online Databases." Retrieved 1 February, 2010, from http://web.worldbank.org.ludwig.lub.lu.se/WBSITE/EXTERNAL/DATASTATISTICS/0,,contentMDK:20398986~menuPK:64133163~pagePK:64133150~piPK:64133175~theSitePK:239419,00.html.
6) UNDP. (2008). "Calculating the human development indices." Retrieved 18 February, 2010, from http://hdr.undp.org/en/media/HDR_20072008_Tech_Note_1.pdf.
7) UNDP. (2010). "Human Development Reports." Retrieved 19 February, 2010, from http://hdr.undp.org/en/statistics/data/.
8) UNDP. (2010). "Is the HDI available before 1975?" Retrieved 23 May, 2010, from http://hdr.undp.org/en/statistics/faq/question,78,en.html.
9) UNDP. (2010). "What is the human development index (HDI)?" Retrieved 18 February, 2010, from http://hdr.undp.org/en/statistics/indices/hdi/question,68,en.html.
10) Ura, K. (2008). "Explanation of GNH Index." Retrieved 5 May, 2010, from http://www.grossnationalhappiness.com/gnhIndex/intruductionGNH.aspx.
11) Wikipedia. (2010). "Financial crisis." Retrieved 11 May, 2010, from http://en.wikipedia.org/wiki/Financial_crisis.

# Appendix

## A1 Maximizing the welfare function

The welfare maximization problem can be formulated as follows if there is a total amount $X^1, \dots, X^k$ of goods $1, \dots, k$ to distribute among the consumers, Equation 10:

$$\begin{aligned} & max\ W\big(u_1(x), \dots, u_n(x)\big) \\ & such\ that \sum_{i=1}^{n} x_i^1 = X^1 \\ & \vdots \\ & \sum_{i=1}^{n} x_i^k = X^k \end{aligned}$$

**Equation 10**

The maximal welfare allocation must be a Pareto efficient allocation because if it were not then there would be some other feasible allocation that gave everyone at least as large a utility, and someone strictly greater utility. The figure below, Figure 19, illustrates the situation where the set U indicates the set of possible utilities in the case of two individuals; known as the utility possibilities set. The utility possibilities frontier is the boundary of this set associated with Pareto efficient allocations. The welfare maximum is on the boundary of the utility possibilities set and there are no other feasible allocations that yield higher utilities for both agents. (Varian 2006)

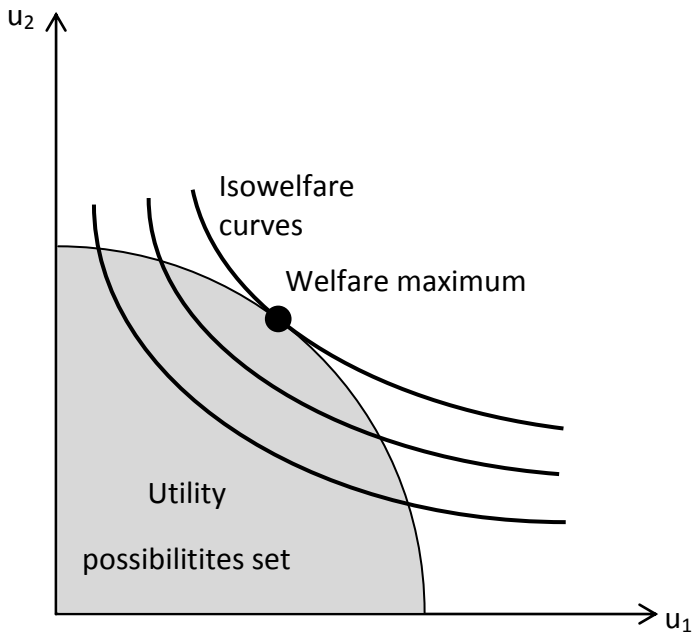



**Figure 19. Welfare maximization. An allocation that maximizes a welfare function must be Pareto efficient**

The Isowelfare curves in the figure depict those distributions of utility that have constant welfare. For the weighted-sum-of-utilities welfare function every Pareto efficient point is a maximum if the utility possibility set is convex. (Varian 2006)

## A2 Principal Component Analysis: finding the components

The different principal components are acquired from the singular value decomposition of the data table $X$. Specifically with $X = P\Delta Q^T$, the matrix of factor scores, denoted F is obtained as, Equation 11:

$$F = P\Delta \qquad \textbf{Equation 11}$$

where $P$ is the $I \times L$ matrix of left singular vectors, $Q$ is the $J \times L$ matrix of the right singular vectors and $\Delta$ is the diagonal matrix of singular vectors. The squared diagonal matrix $(\Delta^2)$ is equal to $\Lambda$ which is the diagonal matrix of the (non-zero) eigenvalues of $X^TX$ and $XX^T$.

The inertia of a column is defined as the sum of the squared elements of this column and is computed as, Equation 12:

$$\gamma_j^2 = \sum_i^I x_{i,j}^2 \qquad \textbf{Equation 12}$$

The sum of all the $\gamma_j^2$ is denoted $I$ and it is called the inertia of the data table or the total inertia. Note that the total inertia is also equal to the sum of the squared singular values of the data table.

The matrix Q gives the coefficients of the linear combinations used to compute the factors scores. This matrix can also be interpreted as a projection matrix because multiplying X by Q gives the values of the projections of the observations on the principal components, Equation 13:

$$F = P\Delta = P\Delta QQ^T = XQ \qquad \textbf{Equation 13}$$

(Abdi and Williams 2010)

## A3 Standardization of Data

To be able to compare the different data standardization is needed. First all of the data needs to be in the form of e.g. per capita or as a % of GDP etc. Fortunately this is already a fact with the gathered statistics from The World Bank, (TheWorldBank 2010). Then the data is in need of more standardization by data scaling to make the different data set comparable. A common method for data scaling occurs after mean centering. Each variable is also divided by its standard deviation, Equation 14. (Brereton 2003)

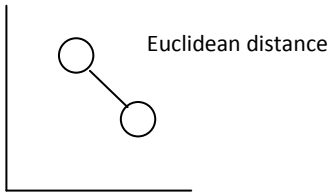
$$x_{ij}^{stn} = \frac{x_{ij} - \overline{x_j}}{\sqrt{\sum_{i=1}^{I} \frac{\left(x_{ij} - \overline{x_j}\right)^2}{I}}} \qquad \textbf{Equation 14}$$

## A4 Cluster Analysis/Dendrogram

The cluster analysis starts off with the search for similarity between the different objects. In general the four most commonly used ways of determining how similar objects are the following (Brereton 2003):

1. Correlation coefficient between samples. A coefficient of 1 means that the samples have identical characteristics and therefore two objects are more similar with a higher correlation coefficient (maximum is however always 1). The correlation coefficient may vary between -1 to +1.
2. Euclidean distance is the distance between two samples *k* and *l* and is defined by, Equation 15:

$$d_{kl} = \sqrt{\sum_{j=1}^{J} (x_{kj} - x_{lj})^2} \qquad \textbf{Equation 15}$$

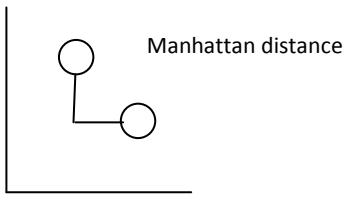



where there are j measurements and $x_{ij}$ is the *j*th measurement on sample i. The smaller the value the more similar are the samples. There is no limit to the Euclidean distance but it is always positive. It can also be presented in matrix format, Equation 16:

$$d_{kl} = \sqrt{(x_k - x_l) * (x_k - x_l)'} \qquad \textbf{Equation 16}$$

3. Manhattan distance is slightly different than the Euclidean distance and is defined thusly, Equation 17:

$$d_{kl} = \sum_{j=1}^{J} |x_{kj} - x_{lj}| \qquad \textbf{Equation 17}$$




The values of the Manhattan distance will always be greater than those of the Euclidean distance.

4. Mahalnobis distance is the last of the four and is a bit more difficult to grasp at the first look. It is defined similar to the Euclidean distance but it also takes into account that some of the variables may be correlated and so measure more or less the same properties. In matrix form this distance is defined, :

$$d_{kl} = \sqrt{(x_k - x_l) * C^{-1}(x_k - x_l)'} \quad \textbf{Equation 18}$$

where C is the variance-covariance matrix of the variables. The matrix is symmetric about the diagonal where the elements represent the covariance between any two variables of dimensions *J x J*.

After the similarities have been defined it is time to link the objects. I seems reasonable to cluster single objects gradually until they form groups a so called agglomerative clustering. (Brereton 2003)

1. One starts of with finding the two most similar objects in the raw data by for example checking their correlation coefficient.
2. Now these two objects can form a group of the two most similar objects. The same procedure is done for all the objects in the raw data.
3. However the most difficult part is how to connect the objects to groups. This can be done by studying the distance between the objects.
   *a)* *Nearest neighbor.* The shortest distance between the objects.
   *b)* *Furthest neighbor.* The opposite to the nearest neighbor.
   c) *Average linkage.* There are two different ways of doing this and it is the unweighted and the weighted. If group A consists of $N_A$ objects and group B of $N_B$ objects, the new unweighted similarity measure is given by, Equation 19:

$$s_{AB} = \frac{N_A s_A + N_B s_B}{N_A + N_B} \quad \textbf{Equation 19}$$

and the weighted similarity measure is given by Equation 20:

$$s_{AB} = \frac{s_A + s_B}{2} \quad \textbf{Equation 20}$$

When the linkage is completed for one group the same procedure is to be done for all of them until all objects have joined one large group and then a dendrogram can be made. As seen in Figure 20 a dendrogram is shown were the groups 2 and 5 are most similar and object 1 I very different from the others because it is connected last in a entire different place.

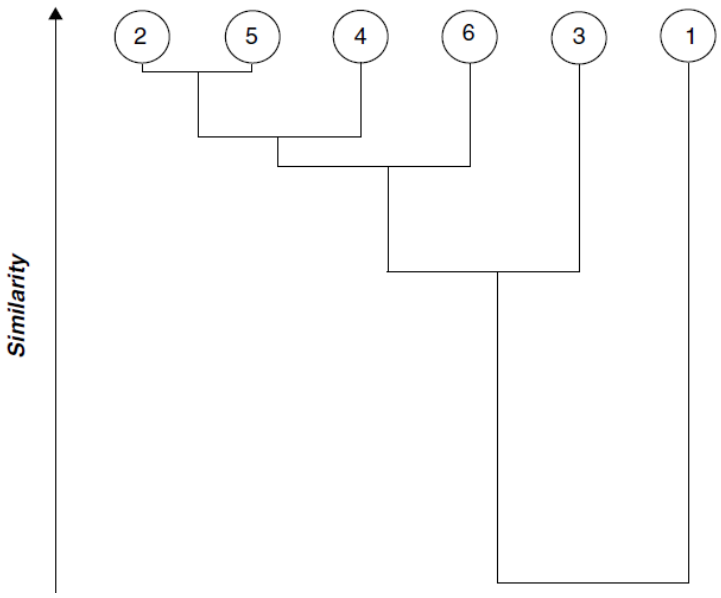



**Figure 20. Dendrogram for cluster analysis example (Brereton 2003)**

# A5 Linear regression

## A5.1 Simple linear regression

Simple linear regression aims to find a straight line trough a set of data points with the sum of squared residuals (vertical distances between the points of the data set and the fitted line) as small as possible. In short the simple linear regression is a least squares estimator of a linear regression model with a single predictor variable, Equation 21:

$$Y_i = \alpha + \beta x_i + \epsilon_i, where\ i = 1,2, \dots, n \qquad \textbf{Equation 21}$$

In this equation $e_i$ are random variables and $\alpha$ and $\beta$ is unknown parameters which describes the linear correlation.

When a line is estimated the sum of least squares a minimized in regard to $\alpha$ and $\beta$, Equation 22:

$$\underbrace{\min}_{\alpha,\beta} S(\alpha,\beta), where\ S(\alpha,\beta) = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 \qquad \textbf{Equation 22}$$

## A5.2 Residuals

In the linear model, Equation 23:

$$Y_i = \alpha + \beta x_i + \epsilon_i \qquad \textbf{Equation 23}$$

it's assumed that $\epsilon_i$ is independent $N(0, \sigma^2)$ with a constant variance $\sigma^2$. To check if these assumptions are somewhat satisfied usually the residuals are examined, Equation 24:

$$r_i = y_i - \alpha^* - \beta^* x_i \qquad \textbf{Equation 24}$$

which are approximations to $\epsilon_i$ and describes the observations distribution around the estimated line. (Olbjer 2000)

### *A5.2.1 Normal distribution*

The standard normal distribution is the simplest cast of normal distribution. This is described by the probability density function, Equation 25 (Olbjer 2000):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty \qquad \textbf{Equation 25}$$

The $\mu$ parameter denotes the median (expectation-value) of the distribution, and $\sigma^2$ is a measure of the width of the distribution (variance). It is not possible to integrate the normal distribution probability density function analytic which is way table often are used when calculating probability. (Olbjer 2000)

The expected value shall correspond to the mean $\bar{x}$ at infinite observations, Equation 26:

$$\mu = \sum (value) * (probaility) = \sum_{k=-\infty}^{\infty} kp(k) \qquad \textbf{Equation 26}$$

The expectation value corresponds to the centre of gravity in the distribution it is assumed that every value is weighted with their corresponding probability. In continuous distributions the expected-value can be written as, Equation 27 (Olbjer 2000):

$$\mu = \int_{-\infty}^{\infty} xf(x)dx \qquad \textbf{Equation 27}$$

The variance $\sigma^2$ for the random variable X, Equation 28 (Olbjer 2000):

$$\sigma^2 = \sum (Outcome - \mu)^2 * (probability) = \sum_{k=-\infty}^{\infty} (k-\mu)^2 p(k) \qquad \textbf{Equation 28}$$

### A.5.3 Heteroscedasticity and homoscedasticity

- "The term Homoskedastic means that the spread of the y variable remains constant as the x values change." (Barreto and Howland 2006)
- "Heteroskedasticity means that the spread of y varies as x changes." (Barreto and Howland 2006)

### A.5.4 Engle's Autoregressive conditional heteroskedasticity (ARCH) model

ARCH is a statistical method to estimate the conditional mean and variance of dataset. The ARCH model can be formulated in terms of an data set, $\psi_t$, over time, $t$. If $y_t$ is the dependent variable and $x_t{}'$ is a vector of explanatory variables included in $\psi_{t-1}$ the $p$th order linear ARCH regression model can be formulated as follows, Equation 29 (Engle 1983):

$$y_t|\psi_{t-1} \sim N(x_t\,\beta, h_t)$$
$$h_t = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \alpha_2\epsilon_{t-2}^2 + \cdots + \alpha_p\epsilon_{t-p}^2 \qquad \textbf{Equation 29}$$
$$\epsilon_t = y_t - x_t\beta$$

The conditional distribution of $y_t$ is normal but the ys are not jointly or marginally normal. The joint density is given as the produkt of all the conditional densitites and the log likelihood function can be formulated as follows, Equation 30 (Engle 1983):

$$L = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log(h_t) - \frac{1}{2}\sum_{t=1}^{T}\frac{\epsilon_t^2}{h_t} \qquad \textbf{Equation 30}$$

The two sets of paramters, $\alpha$ and $\beta$ are prooven to be asymptotically indepentend and therefore the likelihood can be maximized separately for each. In order for the model to work properly ther are some restrictions on the adimissible coefficients. All the alpha parameters are restricted to be nonnegative because if these coefficients are negativ single large residual could drive the conditional variance negative. The process will have an infinite variance of the alphas are too large which makes a second restriction restrict the alphas to lie in a stable region. This is done by the following differens equation which much be stable, Equation 31 (Engle 1983):

$$z_t = \alpha_1 z_{t-1} + \cdots + \alpha_p z_{t-1} \qquad \textbf{Equation 31}$$

With the condition that the sum of all alpha must be less than unity.

As in most time series model the likelihood function above is conditional on initial valuse. The initial disturbance are assumed to be drawn from the unconditional distribution when starting the algorithm. This is only an approximate eighth-order lag but any errors should die out very rapidly and are not responsible for the results reported here as can be shown by using smaller values for p (Engle 1983).

Before estimating the ARCH model the following hypothesis test should been performed because the model is simply estimated by OLS under the null, a hypothesis test based upon the Lagrange multiplier principle is very simple and retains all the asymptotic properties of likelihood ratio tests, Equation 32 (Engle 1983):

$$\alpha_1 = \cdots = \alpha_p \qquad \textbf{Equation 32}$$

## A6 The chosen variables

These are the author's argument for the chosen variables in this essay and the focus is on the welfare of the individual:

## A6.1 Economic variables

Three economic variables have been chosen for the calculations, Table 9:

**Table 9. Economic variables**

| Economic variables | The economy | The welfare | The trade |
|---|---|---|---|
| **GDP growth (annual %)** | The growth provides more investment possibilities which can generate more inventions and higher employment rates for the individual. | A higher GDP should provide an increased spending on social services such as healthcare and education. | A growing economy should provide higher employment rates due to the productivity increase which secures income for the people thus gaining a greater possibility to trade. |
| **Inflation, GDP deflator (annual %)** | Players on the market (individuals and companies) know the real | Connected to the economic argument individuals and | The monetary and fiscal policy is more easily managed when the inflation is |

| | | | |
|---|---|---|---|
| | value of money. | companies are more inclined to make investments. | stable which in some way define the trade options. |
| **High-technology exports (% of manufactured exports)** | The advanced technology broadens the possibilities for creation and creativity. | New technology ensures for example better healthcare and leisure activities. | More export is usually tightly connected with trade. |

## A6.2 Socio-economic variables

Eight socio-economic variables have been chosen for the calculations, Table 10:

**Table 10. Socio-economic variables**

| Socio-economic variables | The economy | The welfare | The trade |
|---|---|---|---|
| **Immunization, measles (% of children ages 12-23 months)** | If more children have a healthy childhood they have a greater possibility of providing to the community. | Immunization against deceases decrease the cost for the healthcare system and instead other illnesses can be addressed. A healthy childhood is also a condition for a long life. | A healthy population ensures a continuous workforce and a low and stable stress on the community. The health of a population can reduce the interests of trading. |
| **Internet users (per 100 people)** | Information sharing is a necessity for innovations and new technology. | Several possibilities is provided on the Internet which both can educate and entertain the people. | An open society includes information sharing and the freedom of speech in which the Internet can be a channel. |
| **Life expectancy at birth, total (years)** | A longer life means longer possible service to the community. | Usually a long life is closely connected to a healthy life. | An active and healthy elderly population will both consume and provide more input and output in the community. |
| **Mobile cellular subscriptions (per 100 people)** | Information sharing is a necessity for innovations and new technology. | A mobile telephone ensures the possibility to contact the authorities when in a vulnerable situation. | The mobile telephone market as well as the Internet market can generate several job opportunities and is an agent for communica- |

| | | | |
|---|---|---|---|
| | | | tion which opens up a society. |
| **Population growth (annual %)** | Two heads are better than one when it comes to innovations and the more innovators there are the more innovations can be created. | More players on the market generate a higher competition which drives the innovations to new highs. | A large population ensures a better diverse cohort which minimizes educational and structural gaps in the community. |
| **Roads, paved (% of total roads)** | Infrastructure in all areas is important for productivity including everything from transportation of goods to transportation of people. | Paved roads are a foundation for a working infrastructure with transportation. | Investments in infrastructure are commonly used in times of crisis and is needed for an effective trade both internally and externally. |
| **Mortality rate, under-5 (per 1,000)** | If more children have a healthy childhood they have a greater possibility of providing to the community. | A healthy childhood is a condition for a long life. | A healthy population ensures a continuous workforce and a low and stable stress on the community. |
| **Time required to start a business (days).** | Some innovations needs to be implemented fast, then it's important to reduce the number of bottlenecks in the innovation process. | An easy and fast system for starting new business lets the originators focus on the company and not the administration process. | Several businesses play at the international market with both imports and exports. More business leads to more trade. |

## A6.3 Environmental variables

Three environmental variables have been chosen for the calculations, Table 11:

**Table 11. Environmental variables**

| Environmental variables | The economy | The welfare | The trade |
|---|---|---|---|
| **Renewable internal freshwater resources per capita (cubic meters)** | Renewable freshwater is one important part in the health of the people. | Renewable freshwater is a foundation for good health in a community. | Renewable freshwater is a foundation for a stable community which is needed for the society in general. |
| **Electric power consumption (kWh per capita)** | A high electric power consumption is assumed to be closely connected to advance technology production which is good for the innovations. | Most welfare needs electric power in one or the other way and a higher consumption enables several different institutions. | As well as for the welfare a high consumption of electric power enables a high technology community which promotes trade. |
| **CO2 emissions (metric tons per capita)** | From a productivity point of view a high CO2 emission is connected to a well industrialized country which is the birthplace for several innovations. | Even though a high CO2 emission might influence the health of the community negatively it also is necessary in some extent to ensure a working industry which generates several welfare components. | A prospering industry ensures a stable community with employment and innovations which stimulates trade. |

## A7 Countries

### A7.1 Countries included in the study

The following 115 countries are included, Table 12.

**Table 12. Countries included in the study**

| Countries included in the study | | | |
|---|---|---|---|
| Albania | Denmark | Korea, Rep. | Russian Federation |
| Algeria | Dominican Republic | Kyrgyz Republic | Saudi Arabia |
| Argentina | Ecuador | Latvia | Senegal |
| Armenia | Egypt, Arab Rep. | Lebanon | Singapore |
| Australia | El Salvador | Lithuania | Slovak Republic |
| Austria | Estonia | Luxembourg | Slovenia |
| Azerbaijan | Ethiopia | Macedonia, FYR | South Africa |
| Bahrain | Finland | Malaysia | Spain |
| Bangladesh | France | Mexico | Sri Lanka |
| Belarus | Gabon | Moldova | Sudan |
| Belgium | Georgia | Mongolia | Sweden |
| Benin | Germany | Morocco | Switzerland |
| Bolivia | Ghana | Mozambique | Syrian Arab Republic |
| Bosnia and Herzegovina | Greece | Namibia | Tanzania |
| Botswana | Guatemala | Nepal | Thailand |
| Brazil | Honduras | Netherlands | Togo |
| Brunei Darussalam | Hungary | New Zealand | Trinidad and Tobago |
| Bulgaria | Iceland | Nicaragua | Tunisia |
| Cambodia | India | Nigeria | Turkey |
| Cameroon | Indonesia | Norway | Ukraine |
| Canada | Iran, Islamic Rep. | Oman | United Arab Emirates |
| Chile | Ireland | Pakistan | United Kingdom |
| China | Israel | Panama | United States |
| Colombia | Italy | Paraguay | Uruguay |

| | | | |
|---|---|---|---|
| Costa Rica | Jamaica | Peru | Venezuela, RB |
| Cote d'Ivoire | Japan | Philippines | Vietnam |
| Croatia | Jordan | Poland | Yemen, Rep. |
| Cyprus | Kazakhstan | Portugal | Zambia |
| Czech Republic | Kenya | Romania | |

## A7.2 Countries not included in the study

The following countries (in total 96) were not included in the study due to that not enough statistics were available, Table 13.

**Table 13. Countries not included in the study**

| Countries not included in the study | | |
|---|---|---|
| Afghanistan | Guam | New Caledonia |
| American Samoa | Guinea | Niger |
| Andorra | Guinea-Bissau | Northern Mariana Islands |
| Angola | Guyana | Palau |
| Antigua and Barbuda | Haiti | Papua New Guinea |
| Aruba | Hong Kong SAR, China | Puerto Rico |
| Bahamas, The | Iraq | Qatar |
| Barbados | Isle of Man | Rwanda |
| Belize | Kiribati | Samoa |
| Bermuda | Korea, Dem. Rep. | San Marino |
| Bhutan | Kosovo | Sao Tome and Principe |
| Burkina Faso | Kuwait | Serbia |
| Burundi | Lao PDR | Seychelles |
| Cape Verde | Lesotho | Sierra Leone |
| Cayman Islands | Liberia | Solomon Islands |
| Central African Republic | Libya | Somalia |
| Chad | Liechtenstein | South Asia |
| Channel Islands | Macao SAR, China | St. Kitts and Nevis |
| Comoros | Madagascar | St. Lucia |

| | | |
|---|---|---|
| Congo, Dem. Rep. | Malawi | St. Vincent and the Grenadines |
| Congo, Rep. | Maldives | Suriname |
| Cuba | Mali | Swaziland |
| Djibouti | Malta | Tajikistan |
| Dominica | Marshall Islands | Timor-Leste |
| Equatorial Guinea | Mauritania | Tonga |
| Eritrea | Mauritius | Turkmenistan |
| Faeroe Islands | Mayotte | Uganda |
| Fiji | Micronesia, Fed. Sts. | Uzbekistan |
| French Polynesia | Monaco | Vanuatu |
| Gambia, The | Montenegro | West Bank and Gaza |
| Greenland | Myanmar | Virgin Islands (U.S.) |
| Grenada | Netherlands Antilles | Zimbabwe |

## A8 Dendrogram

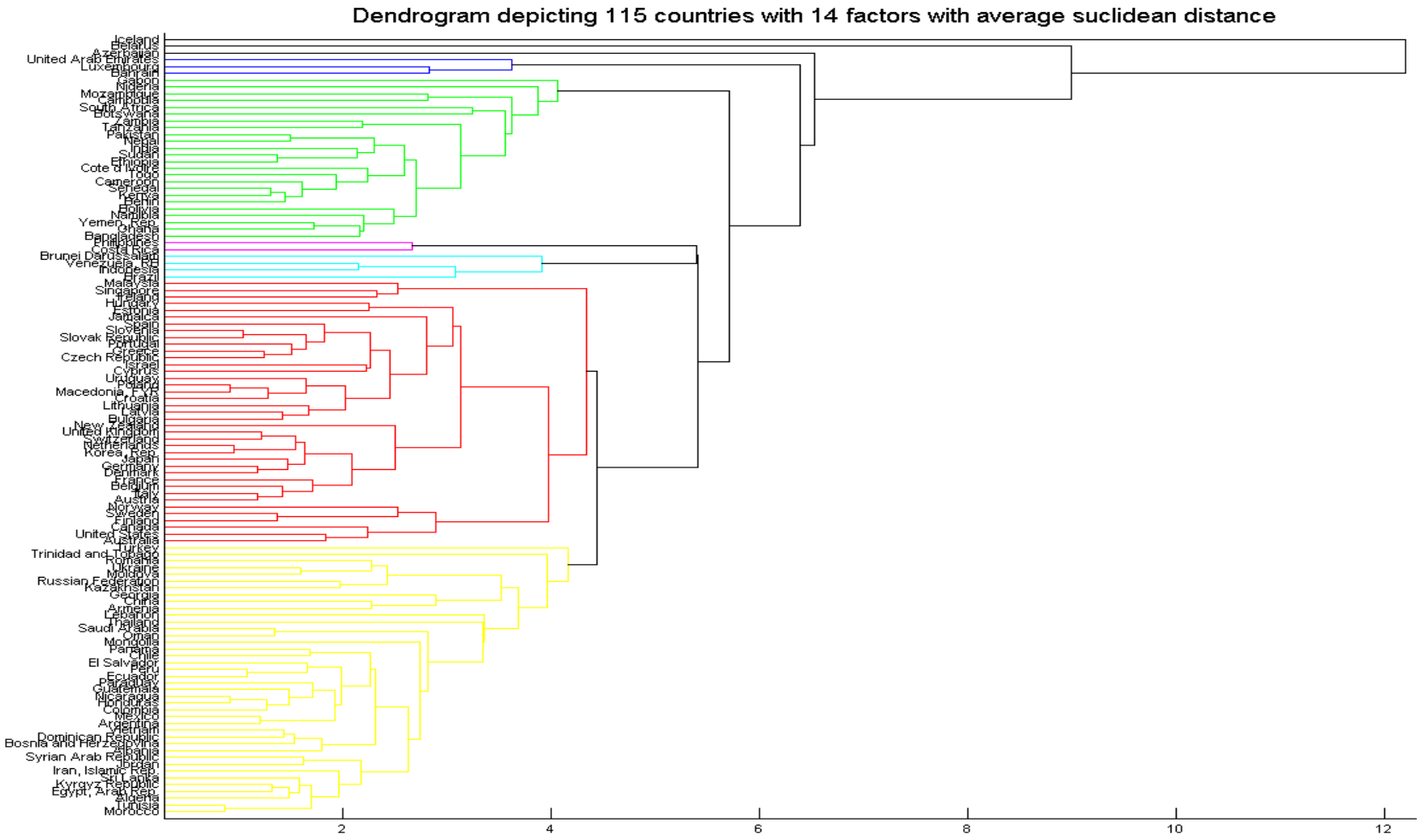



**Figure 21. Dendrogram**

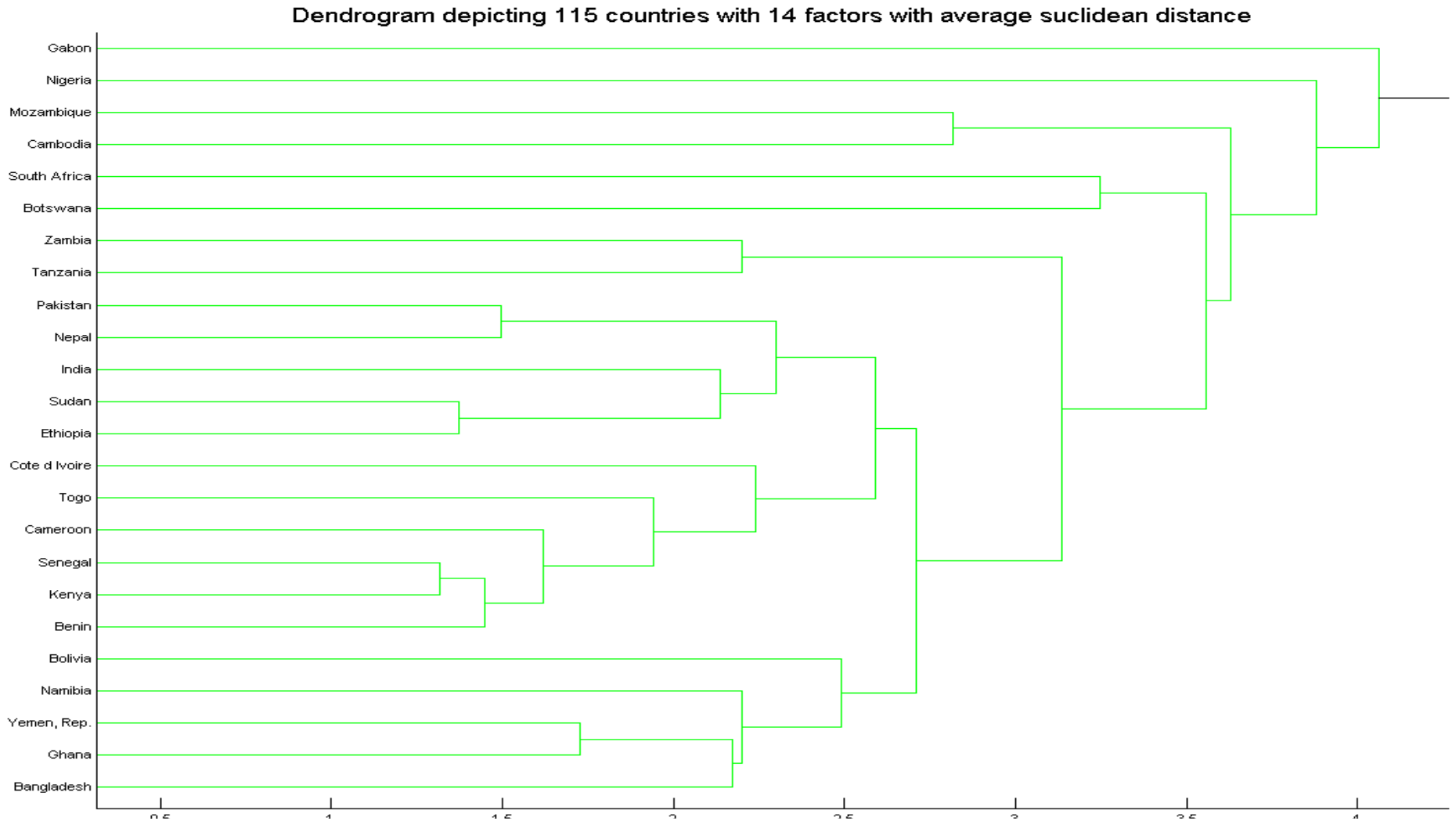



**Figure 22. Cluster 1 of the dendrogram**

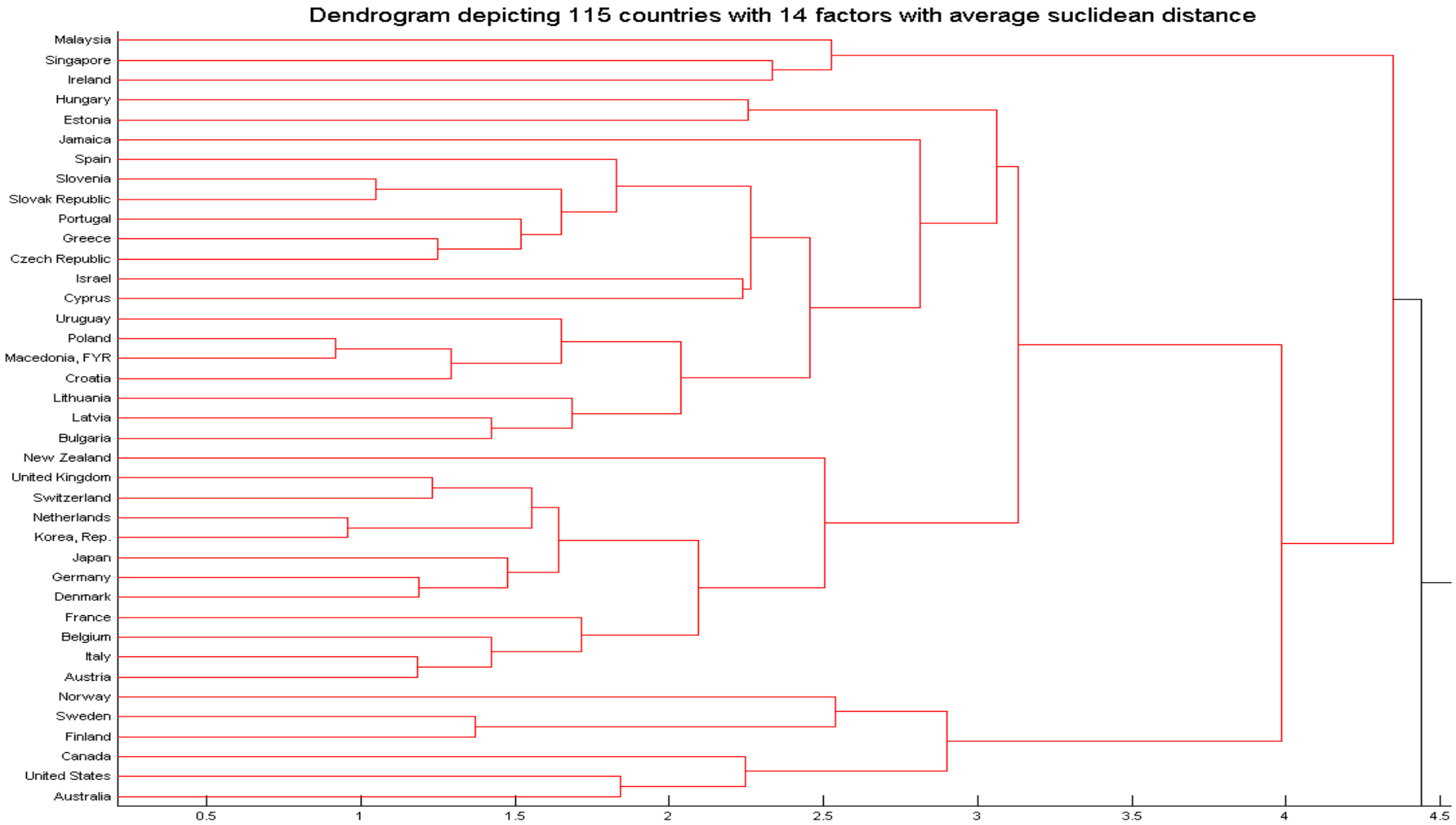



**Figure 23. Cluster 2 of the dendrogram**

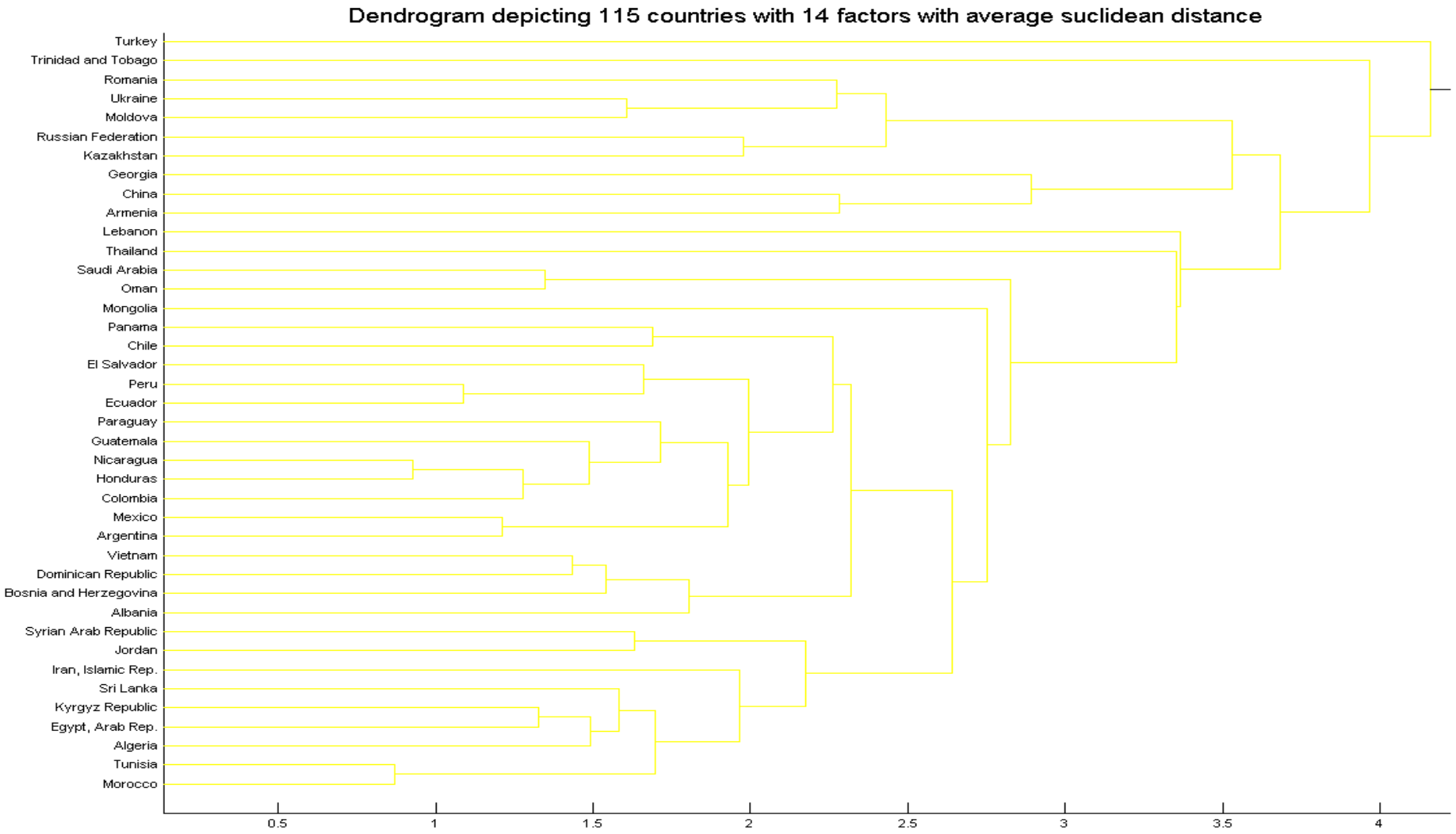



**Figure 24. Cluster 3 of the dendrogram**

## A9 Time series plots

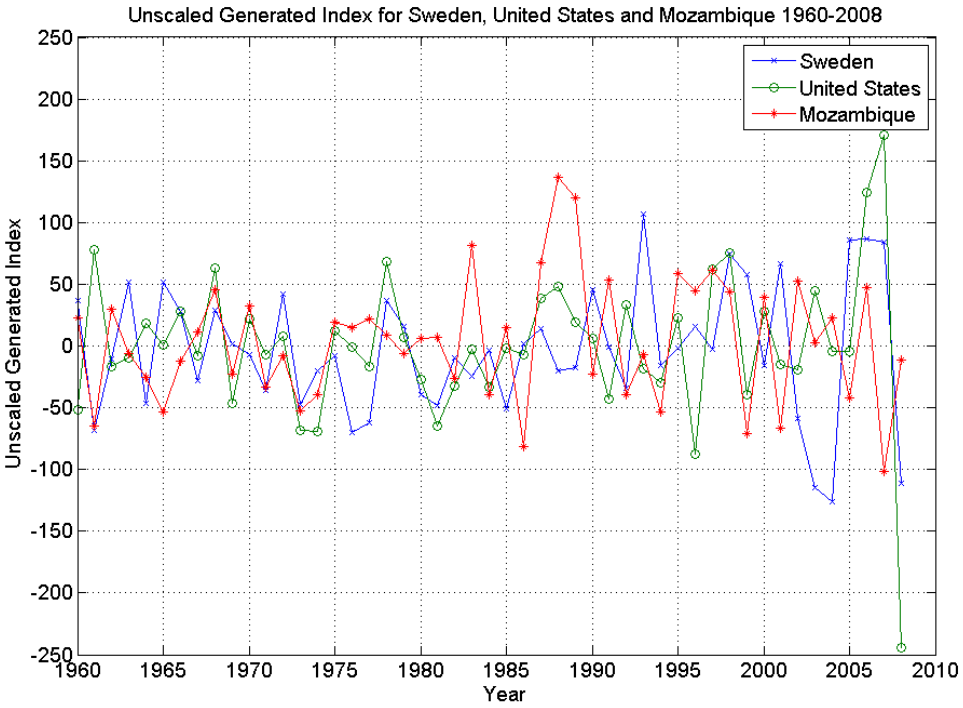



**Figure 25. Unscaled Generated Index for Sweden, United States and Moçambique 1960-2008**

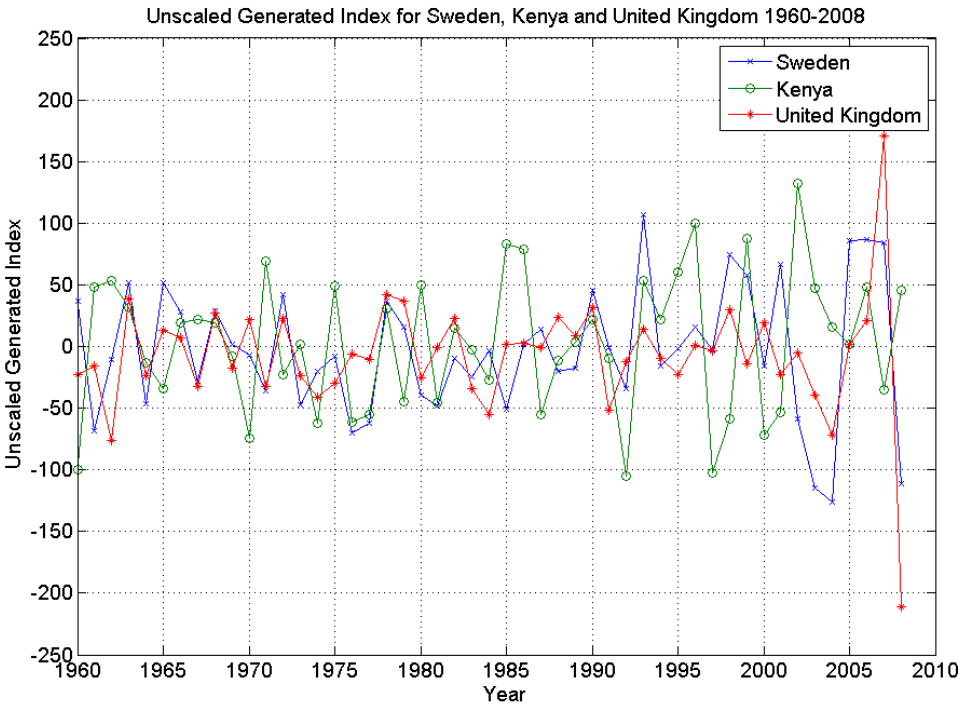



**Figure 26. Unscaled Generated Index for Sweden, Kenya and United Kingdom 1960-2008**

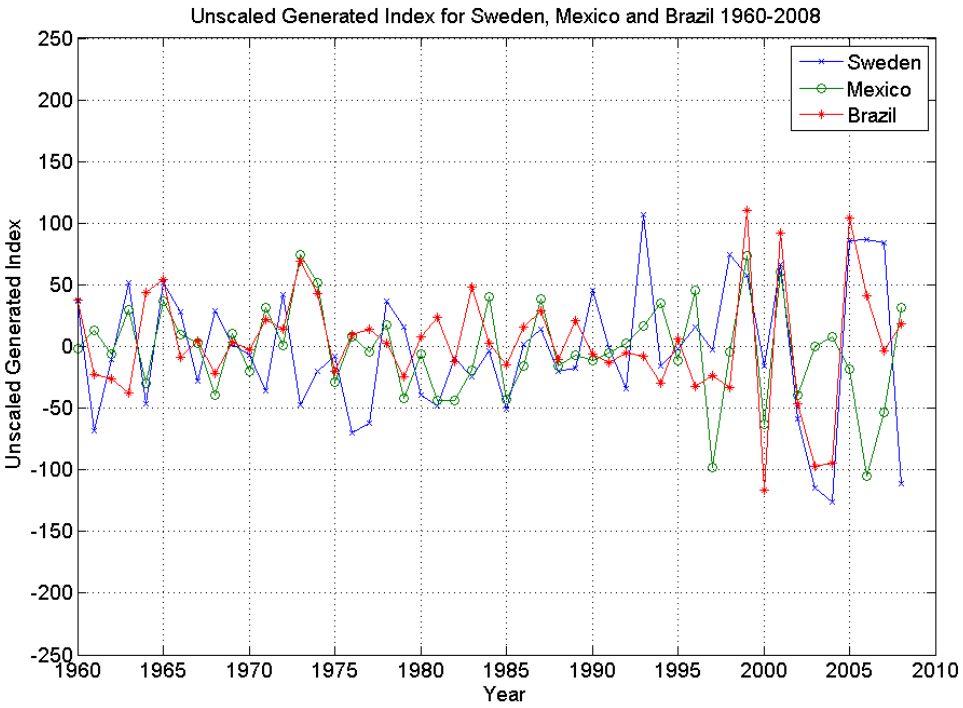



**Figure 27. Unscaled Generated Index for Sweden, Mexico and Brazil 1960-2008**

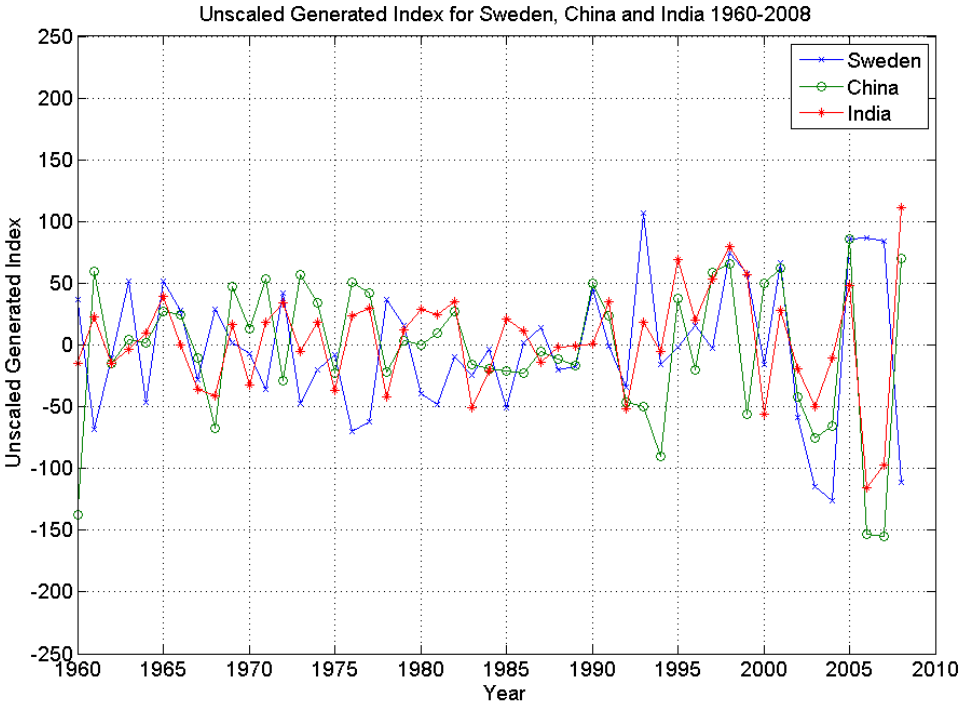



**Figure 28. Unscaled Generated Index for Sweden, China and India 1960-2008**

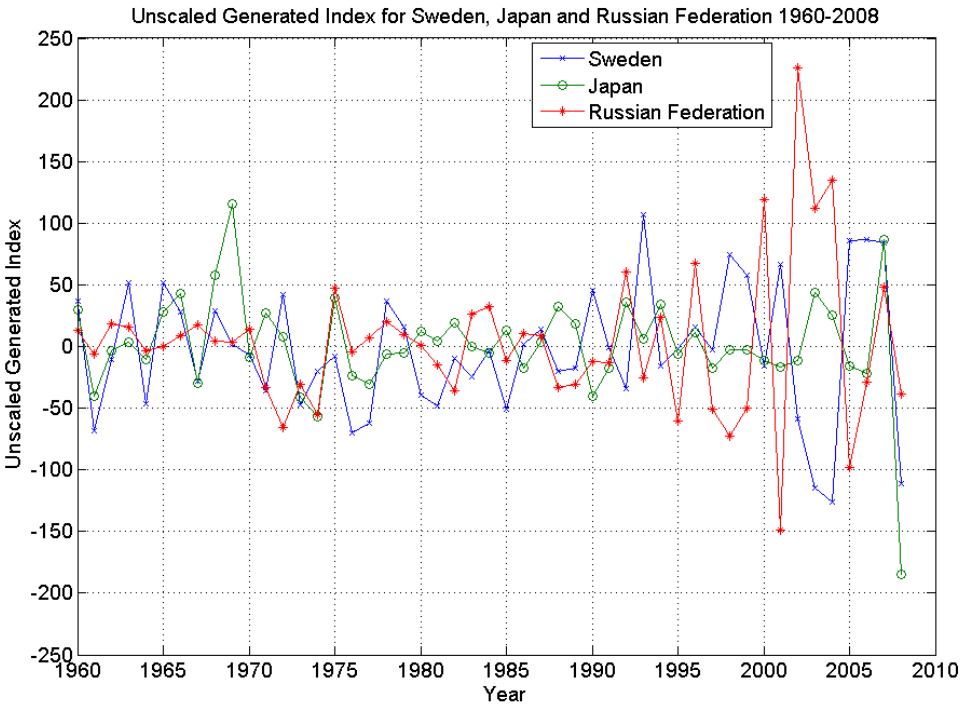



**Figure 29. Unscaled Generated Index for Sweden, Japan and Russian Federation 1960-2008**