EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Institute for the Protection and Security of the Citizen

A New Participatory Framework to Build and Interpret Composite Indicators: An Application to Country Competitiveness

Rabin E.J. Neslo, Roger M. Cooke, Stefano Tarantola

Institute for the Protection and Security of the Citizen

2006

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Statisticians often argue that all the work done during the composition of a composite indicator, to collect data and to edit this data, seems to be "hidden" or "wasted". On the other hand stakeholders like the idea that a composite indicator is capable of summarizing complex and or difficult concepts into one single number.

The research behind this thesis will start by changing the process of obtaining a composite indicator. At first a composite indicator will be obtained by eliciting stakeholders' preferences. After obtaining the composite indicator the work that statisticians put into collecting and editing data, is used to identify the drivers behind the concept of the composite indicators, using regression.

The methods to elicit stakeholders' preferences are not new and are known as paired comparison analysis. However new models are used to obtain an index out of the paired comparison data. These new models make use of Probabilistic Inversion and use IPF or PARFUM to solve them. Also these new models make no assumptions about the distribution of the scores of the alternatives as the traditional methods do, but obtain better variances estimated and covariance.

Keywords : Composite indicators, stakeholders' preferences, regression, paired comparison analysis, Probabilistic Inversion, IPF, and PARFUM.

1. INTRODUCTION

The **Joint Research Center** (JRC) of the **European Commission** is an independent institution that gives scientific and technical support to all Member States of the **European Union**(EU). Research is conducted at the JRC and the results are presented to the European Commission so that the European Commission can adapt or improve policies and decisions.

The research is conducted at the **Institute for the Protection and Security of the Citizen** (IPSC) and more precisely at the Unit, **Econometrics and Statistical Support to Anti-Fraud** with the collaboration of the TU Delft. One of the projects within this Unit is to develop and evaluate composite indicators (indices) of countries performance in many different fields (internal market, e-business, innovation, knowledge economy, etc.).

An index over countries is a numerical scale used to compare countries with one another or with some reference number based on one or more specified aspects. Often an index is used in the decision making of policymakers, because it provides policymakers with information about the monitoring of countries progresses in a given policy field.

There are several other institutions (UNDP, WEF, World Bank, etc.) that elicit indices. When policymakers use an index in their decision making it is important that the index is independent from the institution eliciting it. Besides the index must be a single unambiguous number which captures the facts of the specified aspects and or the different opinions of stakeholders about the specified aspects.

The methods usually applied for eliciting an index often make use of surveys and statistical data. In the surveys respondents are asked to judge the countries based on a specified aspect. The judgment is based on a scale defined by the institutions. The judgments over the specified aspects together with the statistical data are first normalized to uniform scale and then aggregated using an arbitrary set of weights to create an index.

The above method reflects the preference of the institution rather then the preferences of the stakeholders even though the preferences of stakeholders may be captured by the use of surveys. The institutions eliciting an index directly impose their preferences of the several aspects of an country by the use of normalization and the selection of a set of arbitrary weights, because they decide what is important and what is not.

The statistical data used may also be irrelevant in the elicitation of an index. Statistical data become irrelevant if either the data do not reflect any of the specified aspects or if the data only reflect the specified aspects for a subset of countries.

The above discussion initiated the search for alternative methods for developing indices over countries. In this research the Global Competitiveness Index elicited by the World Economic Forum(WEF) is considered as a test case.

Every year the WEF elicits a Global Competitiveness Index (GCI).This index is meant to measure the amount of competitiveness of each country and to shed some light on why some countries grow and others do not in terms of macroeconomics, institution, and technology. Competitiveness is defined as the set of institutions, policies, and factors that determine the level of productivity of a country.

1.1 Objectives of Research

The main objective of this research is to get an index for a set of countries based on preferences of respondents on these countries given the aspect competitiveness. The first phase is to ask respondents to provide us with their preferences on the objects. Initially these respondents will not be actual stakeholders, but are used to test the method of this research. To elicit the preferences of the respondents the Law of Comparative Judgment is used.

In the second phase of this research, the index obtained via the Law of Comparative Judgment is compared with the index from the World Economic Forum to determine if there is any correlation between the two indices.

Finally regression is used to select statistical data that are relevant to the aspect competitiveness of the set of countries. The coefficients obtained from the regression analysis can be seen as the weights selected by the World Economic Forum.

2. BACKGROUND

The World Economic Forum (WEF) claims to be an independent, international organization incorporated as a Swiss not-for-profit foundation. They are striving towards a world-class corporate governance system where values are as important a basis as rules. Their motto is 'entrepreneurship in the global public interest'. They believe that economic progress without social development is not sustainable, while social development without economic progress is not feasible (World Economic Forum — Our Organization[**?**]).

According to its supporters, the World Economic Forum is an ideal place for dialogue and debate regarding the major social and economic problems of the planet, since representatives of both the most powerful economic organizations and the most powerful political organizations are present, since intellectuals also participate, and since there is a generally informal atmosphere encouraging wide-ranging debate. Journalists have access to every session at the Annual Meeting in Davos and the majority of sessions are webcast live so that the debates can be open to a wider public. In all about 600 journalists from print, radio and TV take part in the meeting. Whilst business and political leaders make up the majority of participants, NGO leaders from groups such as Amnesty International, Transparency International, Oxfam and various UN organizations attend, as well as trades union leaders and religious leaders.

According to its critics, the World Economic Forum is really just a business forum, where the richest businesses can easily negotiate deals with one another and lobby the world's most powerful politicians, and that the aim is profit-making rather than solving economic problems like poverty. It has also been criticized as an elitist forum for circumventing democratic politics, and for encouraging non-transparent, secretive decision-making (World Economic Forum — Wikipedia , The Free Encyclopedia[**?**]).

2.1 Support of The WEF

The WEF's membership, the membership of its board, and the attendance at its annual meetings is heavily composed of representatives from Europe, the USA and industrialized Asia.

The main 1,000 member companies are invited to the WEF based on annual revenues of over $1 billion (as of 2002). (This has led some to claim that companies from poorer countries are inherently underrepresented. The WEF claims that 200 companies, mainly from the developing world, are invited to join the WEF membership and events.)

In the 2002 WEF Annual Meeting, 75% of participants were from Europe (39%) and the US (36%), which together represented approximately 17% of the world's population. West Asian participants were about five times overrepresented relative to their population, i.e. they constituted 4% of participants while representing 0.8% of the world's population.

Also, while 60% of the world's population live in Asia (as of 2002), about 7.7% of the participants at the 2002 Annual Meeting were Asian(World Economic Forum — Wikipedia , The Free Encyclopedia[**?**]).

2.2 Related Work of The WEF

One of the WEF related works is the publishing of the Global Competitiveness Report(GCR). The WEF asserts that the GCR has led to the assessment of the competitiveness of nations. According to the WEF the GCR sheds light on the question why some countries are able to grow on a sustained basis for prolonged periods of time, in the process pulling large segments of the population out of poverty, while others remain stagnant or, worse, actually see an erosion of living standards [**?**]. In the GCR the WEF compiles three indices namely, the Growth Competitiveness Index, the Business Competitiveness Index, and the Global Competitiveness Index(GCI).

The focus of this research will be put on the Global Competitiveness Index of the WEF. This index is meant to measure the amount of competitiveness of each country and shed some light why some countries grow and other not in terms of macroeconomics, institution, and technology. Competitiveness is defined as the set of institutions, policies, and factors that determine the level of productivity of a country.

2.2.1 Structure of the GCI

The GCI is the aggregation of nine pillars. These pillars are grouped into three categories namely: basic requirements, efficiency enhancers, and innovation. See the picture below for the grouping.

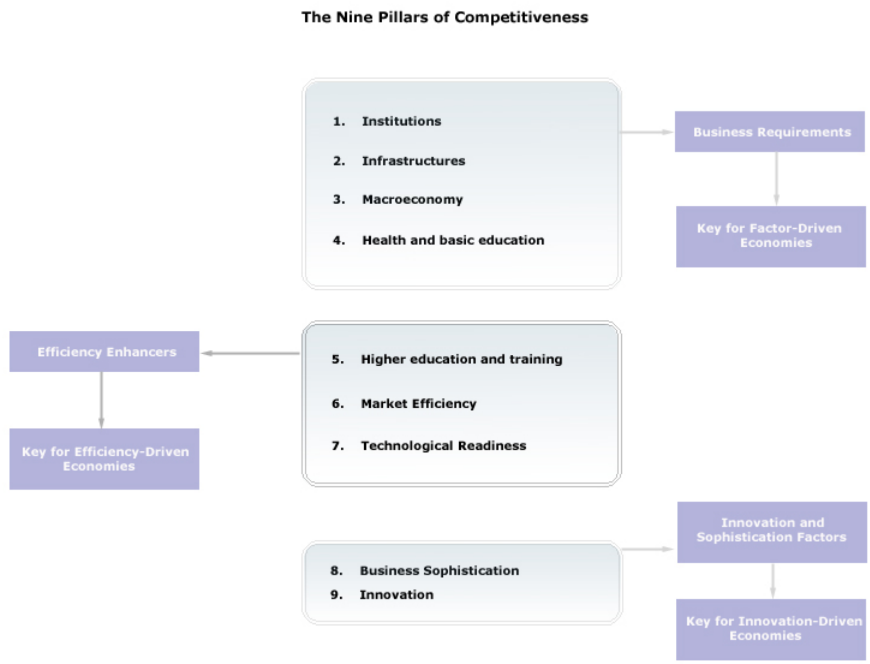

Fig. 2.1: The nine pillars of competitiveness

Each of the pillars is built out of a combination of soft (survey) and hard (statistical) data. For example the pillar Infrastructure is built out of hard data "Telephone lines 2003" and soft data "Overall infrastructure quality", "Railroad infrastructure development", "Port infrastructure quality", "Air transport infrastructure quality", and "Quality of electricity supply"'.

The WEF believes that when compiling the GCI for a country the weights depend on the stage the country is in. A country can be in five different stages namely: stage one (per capita income less than US$2000 or more than 70% of primary exports), transition from stage one to stage two (per capita income between US$2000 and US$3000), stage two (per capita income more than US$3000 and less than US$9000), transition from stage two to stage three (per capita income between US$9000 and US$17000), and stage three (per capita income more than US$17000). Countries that are in the fixed stage from stage one to stage three get fixed weights assigned. Countries that are in the transition stages get weights assigned by the proportion of their income. Each of the pillars is built out of soft (survey) and hard (measured) data. The value of each pillar is an averaged sum of hard and soft data. The hard data is normalized prior to the calculation of the averaged sum mentioned above.

The structure of the GCI with the pillars and data is given by the following picture:

Fig. 2.2: The structure of the Global Competitiveness Index

The WEF method can be formulated as a bottom-up approach. The WEF gathers a set of data that capture different aspects of competitiveness of a country. The approach of the WEF can be visualized with the picture displayed below.

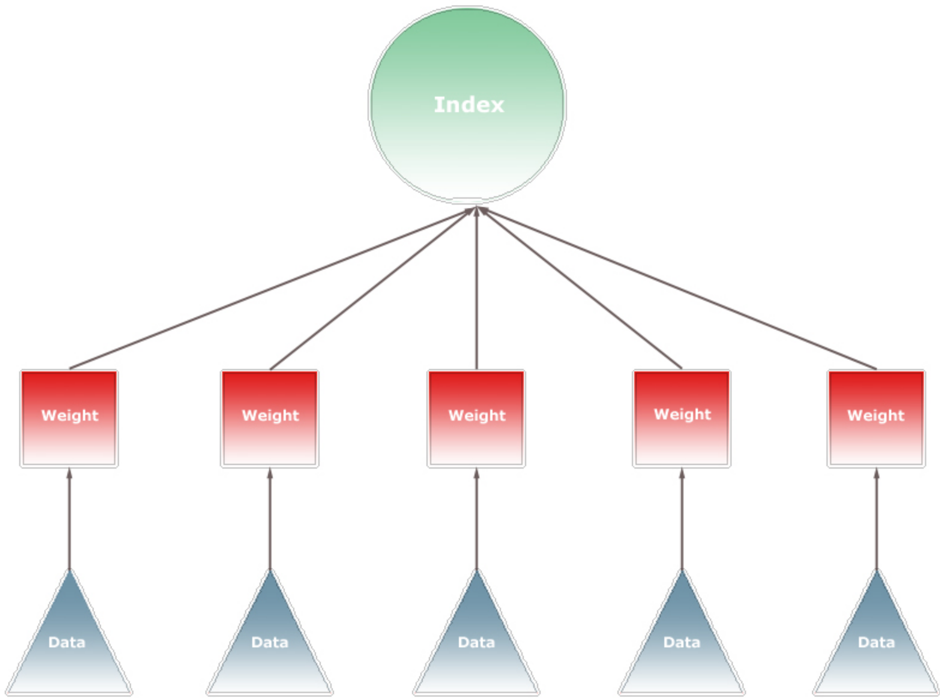

Fig. 2.3: Bottom-Up Approach of The WEF

2.2.2 The Use of Data by The WEF

The WEF uses soft (survey) data and hard (statistical) data in the compilation of the GCI. The soft survey data is obtained using a seven-point Likert scale[1]. After obtaining the responses of the survey the mean over the respondents is used to indicate the score of the survey for each country. The hard data are measured on either an interval level or ratio level. However the WEF transforms the hard data to a one-to-seven scale, to maintain comparability with soft data. This is done by the following formula:

$$\text{Transformed hard data} = 6 * \frac{(\text{original value hard data} - min)}{(max - min)} + 1 \tag{2.1}$$

Max in equation (**??**) is defined as the maximum value of the specific hard data across all the countries and likewise min as the minimum value.

[1] A Likert scale (pronounced 'lick-ert') is a type of psychometric scale often used in questionnaires. It asks respondents to specify their level of agreement to each of a list of statements. It was named after Rensis Likert, who invented the scale in 1932. A typical survey item in a Likert scale is a statement, the respondent is asked to indicate their degree of agreement with the statement

2.3 Summary of The GCI

The WEF tries to identify the variables that influence the level of productivity of a country. The variables are identified by attracting as much as possible the actors that are affecting or that are being affected by the productivity of a country. Although the WEF opens the doors to actors from countries that currently are in a developing stage, most of these actors still come from rich and industrialized countries.

In the construction of the GCI the WEF tries to make a good distinction between the development stages of a country. Basically a development stage of a country can be referred to as the economic level of a country, where economic level is mainly measured in GDP *per capita* [2]. In turn GDP is measured using the expenditure method:

> GDP = *consumption + investment + government spending + (exports - imports)*

The WEF is aware that GDP *per capita* does not give the complete picture of the wealth of a country, because it both does not measure things it should measure or measures things it should not measure. A few examples are:

- economic inequality
- goods on markets which are not observed by statistics agencies i.e. the informal sector
- measurement of non-marketed output like housework or (Do It Yourself) DIY. If a cleaner is hired their income is included in the GDP, but DYI is not.
- externality effects from traded goods, for instance the effects of fuel use on global warming
- defensive expenditure to offset the adverse environmental effects of economic growth such as pollution

Before obtaining the GCI for a country the WEF groups the nine pillars into three basic groups namely: basic requirements, efficiency enhancers and innovation factors. The GCI then becomes

$$\begin{aligned} GCI &= \alpha_1 * basic\ requirements \\ &+ \alpha_2 * efficiency\ enhancers \\ &+ \alpha_3 * innovation\ factors \end{aligned} \tag{2.2}$$

where α_1, α_2, α_3 are weights that sum up to one. Countries that are in a developing stage are mostly busy with enhancing and sustaining the basic requirements so α_1 is higher for these countries. Equation (**??**) tries to filter out

[2] Gross Domestic Product (GDP) is defined as the market value of all final goods and services produced within a country in a given period of time.

the the things that the GDP is not suppose to measure and is not measuring.

There are a number of technically questionable points in the construction of the GCI from the WEF. First point that must be addressed is the selection of the weights. The use of the specified weights by the WEF is justified by the stage criteria of a country, but the derived values for these weights are not backed up by any scientific fact. A country that is in the first stage of development has 0.5 as weight for basic requirements, 0.4 for efficiency enhancers and 0.10 for innovation and sophistication factors, where do the values 0.5, 0.4 and 0.1 come from? Why not other plausible values?

The seven response categories represent an ordinal level of measurement. The categories represent an inherent order but the numbers assigned to the categories do not indicate the magnitude of difference between the categories in the way that an interval or ratio scale would. Measuring central tendency (average) is measured as the mean instead of the mode or median. Measuring the mean is mathematically not defined on an ordinal scale, because subtraction and addition is not defined on an ordinal scale.

The second point is the transformation (**??**) of the hard data from the interval or ratio scale to the one-to-seven scale applied to the hard data. With this transformation the magnitude of difference between the values of the hard data is lost.

Finally the grouping of countries in various stages has the limitation that only the countries in the same stage can be compared with each other.

The bottom-up approach (from disaggregated indicators up to the index) formulated by the WEF is based on a conceptual model (the nine pillars)), which could be criticized. In this research, a top-down approach is implemented where competitiveness is given an operational meaning by stating that one would prefer investing his or her money in a country that is more competitive.

3. THEORETICAL FRAMEWORK

A top-down approach was proposed in the previous chapter to elicit the Global Competitiveness Index (GCI) from people's preferences whether or not to invest in a given country. Based on the definition of the World Economic Forum (WEF) global competitiveness affects people globally. People will form preferences for their country as well for other countries that affect them based on the concept competitiveness. There will be people who prefer a country that has good environmental policies and other who prefer a political stable country. Therefore people preferences on investment are implicitly an aggregated indicator of global competitiveness.

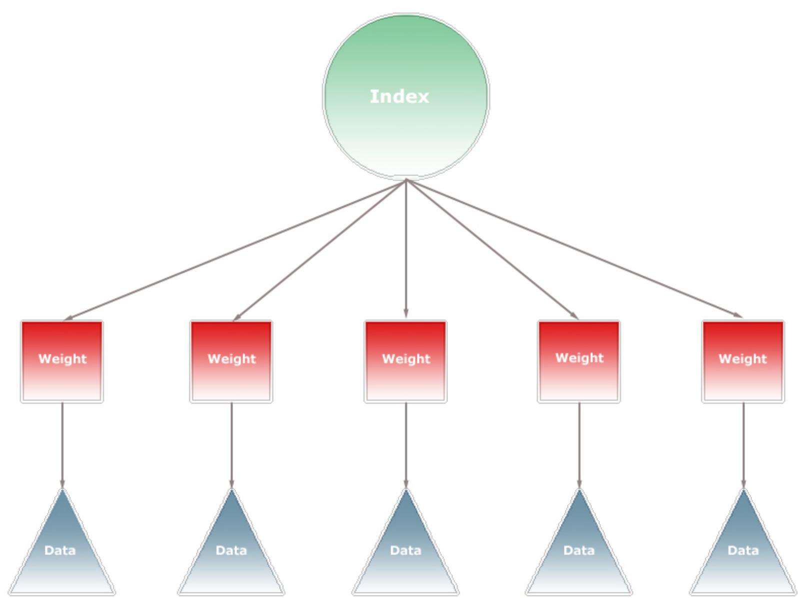

Fig. 3.1: Proposed Top-Down Approach

The preferences will function as a subjective measure for measuring global competitiveness, so the individual preferences need to be aggregated to form a collective preference. In general a group does not behave like a rational individual (Cooke [?]). First it must be shown what a rational individual is and that such an individual has an affine unique subjective utility function over a set of alternatives. This subjective utility function maps from the set of alternatives to the set of real numbers $\mathbb{R}$. So in the case of global competitiveness the subjective utility function of a single rational individual correctly measures the set of countries on a interval scale. After showing how to measure the alternatives for an individual there will be shown how to measure the alternatives for a group of individuals using methods of "Ranked Pairs".

3.1 A Rational Individual

The theory of a rational individual is described in social sciences, especially in economics and by Savage. Savage formulates a more detailed theory of a rational individual, but it does not mean that the one formulated in economics is not correct. The theory of a rational individual described in economics will be briefly reviewed as well as the theory of a rational individual according to Savage.

3.1.1 Thinking Economics

Given a finite set of alternatives or goods a rational individual is one that for each pair of alternatives prefers either one alternative of the pair or both. Further if the individual prefers alternative a over alternative b and alternative b over alternative c then the individual must prefer alternative a over alternative c.

The preferences of an individual is modeled using a preference relation $\preceq$, where $\preceq$ is a binary relation [1]. Let S be the set of alternatives $\preceq$ is a preference relation on S such that if $a \preceq b$ then b is at least as preferable as a for each $a,\ b\ \in S$, b is said to be weakly preferred to a. If $a \preceq b$ but not $b \preceq a$, then the individual strictly prefers b to a, which is written $a \prec b$. If $a \preceq b$ and $b \preceq a$ then the individual is indifferent between a and b, which is written $a \sim b$. A rational individual has a rational preference relation $\preceq$ if is satisfies the following assumptions:

- The relation is reflexive: $a \preceq a$.
- The relation is transitive: $a \preceq b$ and $b \preceq c$ then $a \preceq c$.
- The relation is complete: for all a and b in S we have $a \preceq b$ or $b \preceq a$ or both.

In economics a cunning trader could take advantage of a consumer who has a preference relation that violates transitivity. Suppose the above mentioned consumer has a mango, and prefers mango's to cherries, cherries to passion fruits, and passion fruits to mangos. Then, the consumer would be prepared to

[1] A binary relation (or a dyadic relation) is an arbitrary association of elements of one set with elements of another (perhaps the same) set.

pay, say, one cent to trade their mango for a passion fruit, because they prefer passion fruits to mangos. After that, the consumer would pay once cent to trade their passion fruit for a cherry, and again the cherry for a mango, and so on.

Completeness does not always have to be satisfied. For example a consumer does not have to prefer buying a "Mercedes Benz CLS class" over buying a "BMW sedan 5 series" if he does not have the money for buying either of them. However if the consumer by any chance would win a lottery or is able to dream about it the consumer can prefer either one of the above options or both.

In economics a preference relation on S can be described using a utility function $u : S \rightarrow \mathbb{R}$, such that for every $a,\ b\ \in S$ is $u(a) \leq u(b)$ if and only if $a \preceq b$. A utility function always exists if $\preceq$ is a rational preference relation on S.

3.1.2 Thinking Savage

Di Tillio[?] give the following formulation of Savage's[?] more detailed theory for the preferences of a rational individual.The approach of Savage comes out of the field of decision theory. Savage stated that an individual or subject has his or her belief about a subset of all possible worlds. And that the outcomes of all possible worlds give a certain amount of satisfaction to an individual. The set of all possible worlds is referred to as the *space of uncertainty* Ω and the outcomes as the *space of outcomes* X or consequences, each having an algebra or σ-algebra of subsets. Elements of Ω are referred to as *states of the world* or *states*, and its measurable subsets are called *events*. An individual has a degree of belief for *the family of events* $\mathcal{A}$. This will be referred to as the individual subjective probability.

An *act* is a measurable function of Ω into X, and the set of all acts is denoted by F. A *constant act* is defined as the act that maps every element of Ω into a single element of X. For acts $f,\ g \in X$ and an event $A \in \mathcal{A}$ the act that coincides with f on A and with g on $\Omega \backslash A$ is denoted by $f\ A\ g$. A *preference relation* on F is a binary relation $\preceq$ on F, that is, a subset of $F \times F$. An element $(f, g) \in\ \preceq$ is written $f \preceq g$. If $(f, g) \in\ \preceq$ and $(g, f) \notin\ \preceq$ then $f \prec g$. If $(f, g) \in\ \preceq$ and $(g, f) \in\ \preceq$ then $f \sim g$. Act g is said to be as least as preferable as act f if $f \preceq g$. Act g is strictly preferred over act f if $f \prec g$. One is indifferent over act f and act g if $f \sim g$. At last, an event A is said to be *null* (according to a preference relation $\preceq$) if $f\ A\ h \preceq g\ A\ h$ for all acts $f, g, h \in F$. The subset of $\mathcal{A}$ containing all the non-null events will be denoted by $\mathcal{A}'$.

Di Tillio[?] formulated the following axioms or principles:

P1 (Ordering). *For all acts f and g, either $f \preceq g$ or $g \preceq f$ or both. For all acts f, g, and h if $f \preceq g$ and $g \preceq h$ then $f \preceq h$.*

P2 (Sure-Thing Principle). *Let A be an event, and for all f, g, h, and h'*

acts $\in F$ *. Then* $f\ A\ h \preceq g\ A\ h$ *if and only if* $f\ A\ h^{'} \preceq g\ A\ h^{'}$.

P3 (Event wise Monotonicity). *For each* $A \in \mathcal{A}^{'}$*, let* f *be an act, and let* x *and* y *be outcomes. Then* $x \preceq y$ *if and only if* $xAf \preceq yAf$.

P4 (Weak Comparative Probability). *Let* x*,* y*,* $x^{'}$*, and* $y^{'}$ *be outcomes such that* $x \prec y$ *and* $x^{'} \prec y^{'}$*, and let A,B be events. Then* $xAy \preceq xBy$ *if and only if* $x^{'}Ay^{'} \prec x^{'}By^{'}$.

P5 (Nondegeneracy). *There exist outcomes* x *and* y *such that* $x \prec y$.

P6 (Small Event Continuity). *Let* f *and* g *be acts such that* $f \prec g$*, and let* x *be an outcome. Then there exists a finite, measurable partition* $\left\{A^1, ..., A^N\right\}$ *of such that* $xA^n f \prec g$ *and* $f \prec xA^n g$ *for all* $1 \leq n \leq N$.

UM (Uniform Monotonicity). *Let* f *and* g *be acts such that* $f \prec g$*, let* x *be an outcome, and let* A^1*,* A^2*, ... be events such that* $A^{n+1} \subseteq A^n$ *for all* $n \geq 1$ *and* $\bigcap_{n\geq 1} A^n = \emptyset$*. There exists* $N \geq 1$ *such that* $xA^n f \prec g$ *and* $f \prec xA^n g$ *for all* $n \geq N$.

Savage defines a *utility function* as map $u\ :\ X \rightarrow \mathbb{R}$ and a *subjective belief* that is a finitely additive map $P\ :\ \mathcal{A} \rightarrow [0,1]$ such that $P(\Omega) = 1$. The pair (u, P) is a *subjective expected utility representation* of a preference relation $\preceq$ on F, if $\forall f,\ g\ \in F$

$$f \preceq g \Leftrightarrow \int_\Omega (u \circ f) dP\ \leq \int_\Omega (u \circ g) dP \tag{3.1}$$

The pair (u, P) is *unique* upon a positive affine transformation that is, if there is another pair $\left(u^{'}, P^{'}\right)$ satisfying (**??**), then $u^{'}\ =\ au + b$ and $P = P^{'}$, for some numbers a, b with $a > 0$. A subjective belief P is *convex ranged* if $\forall\ A \in \mathcal{A}$ and every $0 \leq p \leq P(A)$ there exists an event $B \subseteq A$ such that $P(B) = p$. The subjective belief is *dense ranged* if $\forall\ A \in \mathcal{A}$ and every $0 \leq p \leq P(A)$ and every $\epsilon\ >\ 0$ there exists an event $B \subseteq A$ such that $-\epsilon < P(B) - p < \epsilon$. Mover over the subjective belief is *non-atomic* if $\forall\ A \in \mathcal{A}$ such that $P(A) > 0$ there exists an event $B \subseteq A$ such that $0 < P(B) < P(A)$. Finally, if $\mathcal{A}$ is a σ-algebra, then P is a *probability measure* if it is countably additive.

Savage proved that, when $\mathcal{A}$ is the family of all subsets of Ω, axioms or principles P1-P6 imply the existence of a unique subjective expected utility representation with a non constant utility function and a convex ranged subjective belief.

Savage's Theorem. *Assume* $\mathcal{A} \subseteq 2^\Omega$ *is a* σ*-algebra. Then* $\preceq$ *satisfies P1-P6 if and only if there exists a non constant utility function* u *and a convex ranged subjective belief* P *such that* (u, P) *is a subjective expected utility representation* *of $\preceq$. In this case, (u, P) is unique, and P is a probability measure if and only if $\preceq$ satisfies UM* **(Uniform Monotonicity).**

The proof of Savage will be left out in this report, but can be found in[**?**] or in[**?**]. From the principles and theorem of Savage it follows that *a ranking on interval scale can be obtained. The principle of ordering P1 or weak order enables a ranking of a set of alternatives in economics and acts according to Savage. The other principles and the theorem of Savage, or in economics the representation of a preference relation by a utility function, enables the scoring of the alternatives or acts.*

3.2 Group Preferences

A group of individuals can not be treated as a rational individual, even if all the individuals in the group have rational preferences. Moreover a ranking can not be obtained from combining the individual preferences. This can be shown using the following example. Suppose there are three alternatives, A, B and C, and that there are three individuals with preferences as follows:

- Individual 1: C $\preceq$ B $\preceq$ A
- Individual 2: A $\preceq$ C $\preceq$ B
- Individual 3: B $\preceq$ A $\preceq$ C

If C is ranked first, it can be argued that B should ranked first, since two individuals (1 and 2) prefer B to C and only one individual (3) prefers C to B. By the same argument A is preferred to B, and C is preferred to A, by a margin of two to one on each occasion. The three preferences combined together are not transitive and form a cycle.

The above example is known as the voting paradox (also known as Condorcet's paradox or the paradox of voting). It is a situation noted by the Marquis de Condorcet in the late 18th century, in which collective preferences can be cyclic (i.e. not transitive), even if the preferences of individual voters are not. This is paradoxical, because it means that majority wishes can be in conflict with each other. When this occurs, it is because the conflicting majorities are each made up of different groups of individuals.

Condorcet formulated a method that can be referred to as a voting system[2]. In the Condorcet method respondents are asked to rank a set of alternatives.

[2] A voting system is a means of choosing between a number of options, based on the input of a number of voters. Voting is perhaps best known for its use in elections, where political candidates are selected for public office. Voting can also be used to award prizes, to select between different plans of action, or by a computer program to determine the best solution to a complex problem. A voting system consists of the rules for how voters express their desires, and how these desires are aggregated to yield a final result

From these preferences a sum matrix can be obtained where cell i, j means the number of times i is been preferred over j. The row sums are the scores for each alternative and the alternative with the highest score is ranked first.

It might occur that the scores for all the alternatives are the same. When this occurs the situation is known as a 'majority rule cycle', 'circular ambiguity' or 'circular tie'.

Arrow [**?**] formulated a theory that no voting system can possibly meet a certain set of reasonable criteria when there are three or more options to choose from. These criteria are called *unrestricted domain*, *non-imposition*, *non-dictatorship*, *monotonicity*, and *independence of irrelevant alternatives*, and are defined below.

- **unrestricted domain or universality**: the vote must have a result that ranks all possible choices relative to one another, the voting mechanism must be able to process all possible sets of voter preferences, and it should consistently give the same result for the same profile of votes no randomness is allowed in the process.
- **non-imposition or citizen sovereignty**: every possible societal preference order should be achievable by some set of individual preference orders.
- **non-dictatorship**: a special individual should not impose his or hers preference to all the preferences of the others.
- **independence of irrelevant alternatives**: changes in individuals' rankings of irrelevant alternatives (ones outside the subset) should have no impact on the societal ranking of the relevant subset.

Arrow's theorem says that if the decision-making body has at least two members and at least three options to decide among, then it is impossible to meet all these conditions simultaneously.

Often not all the criteria's of Arrows's theorem need to be met at once, because the interest of this research is not of finding an agreement on a preference order among several different alternatives. In other words a group of individuals does not have to be treated as a rational individual. The only criteria's that need to be met in the case of this research are **non-imposition or citizen sovereignty** (all possible preference orderings are allowed), **non-dictatorship** (if the majority in a group of individuals ranks an alternative first then that alternative should be ranked first).

Before analyzing group preferences the first question that must be addressed is, "Is there a significant difference in the alternatives with respect to preference?". First for each individual in the group there must be determined how "rational" the individual is i.e., determine if there significant difference in the alternatives with respect to preference. Finally what is the degree of agreement of the individuals as a whole, because total agreement of the group will violate the criteria **non-dictatorship**.

The methods that can establish a scored ranking out of group preferences are the methods of "Ranked Pairs". The idea behind these methods is to pairwise compare all given alternatives. This can be done by either letting individual's pairwise compare the alternatives directly or by letting them rank order the alternatives and then obtain ranked pairs. The advantage of the former is that intransitivity's can be detected in an individual preference when strict preferences are assumed.

In this research strict preference is assumed and individuals are asked to pairwise compare the alternatives. There are a number of methods that analyze the pairwise comparisons. Basically the idea behind these methods is to fit the data acquired to a probability distribution. For every alternative i, j in the set of alternatives, alternative i preferred over alternative j is given as $P(i \succ j) = \frac{number\ of\ times\ i \succ j}{number\ of\ individuals}$. The fitting is done by either using parametric or non-parametric models. Both type of models will be used for the purpose of this research

3.2.1 Significant Difference in Preference for One Individual

For some individuals in the group it might occur that their perceived values of the alternative are barely distinguishable, which may lead to intransitivity's in the preferences of the alternatives. Intransitivity's might also occur when the number of pairwise comparisons is high, because there choice preference will tend to behave like a random process.

Intransitivity's in an individual preference can be detected as follows. Suppose an individual prefers alternative i to alternative j and alternative j to alternative k then intransitivity will occur if this individual prefers alternative k to alternative i. In this research such events will be referred to as *circular triads*. All intransitivity's in the preferences of the alternatives can be reduced to circular triads (Cooke[**?**]). If the number of *circular triads* increases the individual will less likely be rational.

David[**?**] formulated a hypothesis test for detecting whether an individual is choosing at random. David shows that the number of *circular triads* is given by the following formula

$$C(s) = \frac{n\left(n^2-1\right)}{24} - \frac{1}{2}\sum_{i=1}^{n}\left(\pi_{i,s} - \frac{1}{2}(n-1)\right)^2 \tag{3.2}$$

where s is an individual in the group of individuals, $\pi_{i,s}$ is the number of times that individual s prefers alternative i to some other alternative, and n the number of alternatives. Kendall[**?**] has calculated the probabilities for various values of $C(s)$ given two to ten alternatives under the assumption of random preference. An individual s should be dropped out of the group of individuals, if the number of observed $C(s)$ exceeds $C(s)_{max}$. Where $C(s)_{max}$ is the maximum number of *circular triads* allowed. Kendall shows that for more than seven alternatives the quantity

$$\tilde{C}(s) = v + \left(\frac{8}{n-4}\right)\left(\binom{n}{4}\binom{n}{3} - C(s) + \frac{1}{2}\right) \tag{3.3}$$

with

$$v = \frac{n(n-1)(n-2)}{(n-4)^2}$$

is approximately chi square distributed with v degrees of freedom.

If the random preference hypothesis can not be rejected at the 5% level on the basis of the preference data, then one should consider dropping this individual from the study Cooke[**?**].

3.2.2 Significant Difference in Preference for the Group

It is assumed that there is significant difference in the preferences of each individual in the group. There are two tests to determine the agreement namely the test determining the "Coefficient of Agreement" ϑ and the test determining the "Coefficient of Concordance" ω.

Coefficient of Agreement ϑ

The *coefficient of agreement* ϑ is defined by Kendall as

$$\vartheta = \frac{2\sum\limits_{i\neq j}\binom{\pi_{ij}}{2}}{\binom{n}{2}\binom{S}{2}} - 1 \tag{3.4}$$

where the summation runs over $n(n-1)$ terms, π_{ij} is the number of times alternative i is preferred to alternative j, and S the number of individuals. If ϑ is one there is complete agreement.

The distribution of ϑ can be obtained under the hypothesis that all the agreements of the individuals are due to chance. Kendall[**?**] tabulates the distribution of $\sum\limits_{i\neq j}\binom{\pi_{ij}}{2}$ for small values of S and n (see table B.5, Appendix B of the book of Cooke[**?**]). Kendall argued that for large values of S and n the quantity

$$\vartheta' = \frac{4\left[\sum_{i\neq j}\binom{\pi_{ij}}{2} - \binom{n}{2}\binom{S}{2}\frac{S-3}{2(S-2)}\right]}{S-2} \tag{3.5}$$

is approximately chi-square distributed with

$$\frac{\binom{n}{2}S\left(S-1\right)}{\left(S-2\right)^2} \tag{3.6}$$

degrees of freedom . The hypothesis that all the agreements are due to chance should be rejected at the given significance level.

Coefficient of Concordance ω

The *coefficient of concordance* ω is defined as

$$\omega = \frac{\sum_i\left(R_i - \frac{1}{n}\sum_j R_j\right)^2}{\frac{1}{12}S^2\left(n^3-n\right)} \tag{3.7}$$

where R_i is the sum of ranks calculated by

$$R_i = \sum_{s=1}^{S} R_{i,s}$$

with $R_{i,s}$ the rank of alternative i obtained from responses of individual s. The value of $R_{i,s}$ ranges from one to n. Like the coefficient of agreement a value of one for ω means complete agreement.

For the null hypothesis that the preferences are at random, Siegel[**?**] presents a table of critical values of $\sum_i\left(R_i - \frac{1}{n}\sum_j R_j\right)^2$, for n between 3 and 7 and S between 3 and 20. For n larger than 7,

$$\omega' = \frac{\sum_i\left(R_i - \frac{1}{n}\sum_j R_j\right)^2}{\frac{1}{12}Sn\left(n+1\right)} \tag{3.8}$$

is approximately chi-square distributed with $n-1$ degrees of freedom (Siegel[**?**]).

3.3 Parametric Models

The well known and widely used parametric models are the Thurstone or probit model and Bradley-Terry or logit model. The Thurstone and Bradley-Terry models where developed in the field of psychology. Simultaneously the probit and logit model where developed in the social science field.

3.3.1 Thurstone Model

Thurstone formulated a "Law of Comparative Judgment" to elicit a scaling of alternatives referred by him as stimuli based on the preferences of one or more individuals. The law formulates a model that is not solvable, but Thurstone formulated several cases, which are solvable with some assumptions. Thurstone assumes that the value $V(i)$ of a stimulus i is normally distributed with mean μ_i and standard deviation σ_i. The difference between two stimuli i and j is again normally distributed with mean $\mu_{ij} = \mu_i - \mu_j$ and standard deviation $\sigma_{ij} = \sqrt{\sigma_i^2 + \sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j}$. Where ρ_{ij} correlation coefficient of alternative i and j.

The Thurstone cases that are used in this research are the ones that query multiple individuals once. The perceived value of each alternative i is written as $V(i,s)$ where s is an individual out of the group of individuals. In the cases with multiple individuals it is assumed that the value of each alternative i is independent of the individual so $V(i,s) = V(i)$.

Solution of the Thurstone Models

Suppose individual s is drawn randomly from the group. The probability that s prefers alternative i to alternative j is given by:

$$\begin{aligned} P(V(i) - V(j) > 0) &= p\left(\frac{V(i) - V(j) - \mu_{ij}}{\sigma_{ij}} > \frac{-\mu_{ij}}{\sigma_{ij}}\right) \\ &= p\left(X > \frac{-\mu_{ij}}{\sigma_{ij}}\right) \end{aligned} \tag{3.9}$$

where X is a standard normal variable. This probability is estimated by (ij) where (ij) is defined as the number times that alternative i is preferred to alternative j divided by to number of individuals. Since the probability density function of X is symmetric about zero,

$$P\left(X > \frac{-\mu_{ij}}{\sigma_{ij}}\right) = P\left(X < \frac{\mu_{ij}}{\sigma_{ij}}\right) = \Phi\left(\frac{\mu_{ij}}{\sigma_{ij}}\right) \approx (ij) \tag{3.10}$$

where Φ is the cumulative distribution of the standard normal variable. Taking the inverse cumulative distribution of (**??**) and assuming equality gives

$$\mu_i - \mu_j = \Phi^{-1}\left((ij)\right)\sqrt{\sigma_i^2 + \sigma_j^2 - 2\rho_{ij}\sigma_i\sigma_j} \tag{3.11}$$

Equation (**??**), known as the Thurstone Case A, is the complete form of the "Law of Comparative Judgment" and there are $\begin{pmatrix} n \\ 2 \end{pmatrix}$ equations of (**??**) and $\begin{pmatrix} n \\ 2 \end{pmatrix} + 2n$ unknowns. There is no solution for this system of equations for any value of $n > 2$ or if (ij) is equal to one, or zero.

If the correlation between alternatives is assumed constant, which is one the assumptions that can be made, then there are solutions for $n \geq 6$. Equation (**??**) then becomes.

$$\begin{cases} \mu_i - \mu_j = x_{ij}\sqrt{a_i^2 + a_j^2},\ 1 \leq i,j \leq n \\ a_i^2 \equiv \sigma_i^2 - c,\ i \leq n \\ a_j^2 \equiv \sigma_j^2 - c,\ j \leq n \\ c \equiv \rho\sigma_i\sigma_j,\ 1 \leq i,j \leq n \\ x_{ij} \equiv \Phi^{-1}\left((ij)\right),\ 1 \leq i,j \leq n \end{cases} \tag{3.12}$$

The general system of equations (**??**), is not solved in this research. The σ's can only vary from each other if the correlation term ρ becomes sufficiently small, because the correlation term ρ is constant. The next assumption that will be applied to the "Law of Comparative Judgment", known as the Thurstone Case B, is that the correlation is constant and difference between σ_i and σ_j is small $\sigma_i - \sigma_j = \epsilon$. With the next system of equation is obtained.

$$\begin{cases} \mu_i - \mu_j = x_{ij}\sqrt{\frac{1}{2}\left(1-\rho\right)}\left(\sigma_i + \sigma_j\right),\ 1 \leq i,j \leq n \\ x_{ij} \equiv \Phi^{-1}\left((ij)\right),\ 1 \leq i,j \leq n \end{cases} \tag{3.13}$$

The derivation and solution of (**??**) can be found in the book of Warren[**?**]. The last assumption that can be made, which is known as the Thurstone Case C, is that the correlation term ρ is zero and $\epsilon = 0$.

$$\begin{cases} \mu_i - \mu_j = x_{ij}\sqrt{2}\sigma,\ 1 \leq i,j \leq n \\ x_{ij} \equiv \Phi^{-1}\left((ij)\right),\ 1 \leq i,j \leq n \end{cases} \tag{3.14}$$

The system of equations (**??**) can be solved for $n \geq 4$. There will be more equations then unknowns for $n \geq 4$, which will result in an overdetermined system. The overdetermined system can be solved with least squares by finding estimates $\mu_i^{'}$ of μ_i and $\rho^{'}$ of ρ such that the sum of the squared errors

$$SE = \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\mu_i^{'} - \mu_i^{'} - x_{ij}\sigma\sqrt{2} \right)^2 \tag{3.15}$$

is minimal. Differentiating SE with respect to $\mu_i^{'}$ and using the fact that $x_{ij} = -x_{ji}$, and $\left(\mu_i^{'} - \mu_j^{'}\right) = -\left(\mu_j^{'} - \mu_i^{'}\right)$ gives

$$\frac{\partial SE}{\partial \mu_i^{'}} = -2\sum_{j=1}^{n} \left(x_{ij}\sigma\sqrt{2} - \mu_i^{'} + \mu_j^{'} \right) \tag{3.16}$$

The minimum of each partial derivative is reached for $\frac{\partial SE}{\partial \mu_i^{'}} = 0$. Rearranging terms and dividing by n gives

$$\mu_i^{'} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}\sigma\sqrt{2} + \frac{1}{n}\sum_{j=1}^{n} \mu_j^{'} \tag{3.17}$$

The origin of the values is not yet set and it is convenient to set the origin at the mean of the estimated values, so that

$$\frac{1}{n}\sum_{j=1}^{n} \mu_j^{'} = 0 \tag{3.18}$$

$$\begin{cases} \mu_n^{'} = \frac{1}{n}\sum_{j=1}^{n} x_{ij}\sigma\sqrt{2\left(1-\rho\right)}, \; 1 \le i \le n \\ \sigma > 0 \\ \rho \in [-1, 1] \\ x_{ij} \equiv \Phi^{-1}\left(\left(ij\right)\right), \; 1 \le i, j \le n \end{cases} \tag{3.19}$$

The values obtained from this solution are on a interval scale, because there is no absolute zero point.

Goodness of Fit

The squared error SE is minimized, but it still may be too large, which can lead to a bad fit of the data. Mosteller [**?**] formulates a statistical test under the null hypothesis that the data follows the Thurstone case C. It is given as follows

$$\begin{aligned} y\left(ij\right) &= \mu_i^{'} - \mu_j^{'} \\ \Theta\left(ij\right) &= \arcsin\sqrt{\left(ij\right)} \\ \phi\left(ij\right) &= \arcsin\sqrt{\Phi\left(y\left(ij\right)\right)} \end{aligned} \tag{3.20}$$

where Φ is the cumulative standard normal distribution and arcsin is measured in degrees. Mosteller shows that

$$D = \frac{\sum_{i>j} n\left(\Theta\left(ij\right) - \phi\left(ij\right)\right)}{821} \tag{3.21}$$

is approximately chi square distributed with $\left(\begin{array}{c} n \\ 2 \end{array}\right) - n + 1$ degrees of freedom. It is possible to estimate the probability of seeing a deviation at least as great as D. If this probability is less then significance level α the modeling assumption of equal standard deviation and constant correlation should not be maintained, because the model would fit the data badly. On the other hand if the probability is higher then $1 - \alpha$ it would mean that the squared error SE is smaller then expected and the model of equal standard deviation and correlation fits the data better then expected.

Confidence Bounds

It is possible to estimate confidence bounds for parameters of the Thurstone models, via bootstrap. The procedure is given as follows.

Init:
solve model with given data and obtain values for parameters of the model. Define a list L that will store parameter values, a list K that will store the values $V\left(i,s\right)$, number of runs M, and a significance level α .
For:
$k = 1..M$
Run:
Sample values $V\left(i,s\right)^{'}$ for each individual s in the group from the normal distribution with parameters obtained from *Init* and add each value to the list K. Define $\left(ij\right)^{'}$ as $\frac{number\ of\ times\ V(i,s)^{'} > V(j,s)^{'}}{number\ of\ individuals}$. Solve the model using $\left(ij\right)^{'}$ and store the parameter values in L.
Finalize:
Sort the values for each parameter. The confidence bound for each parameter is obtained by selecting the lower bound, which is at the $\frac{1}{2}\alpha M$ index of L and the upper bound, which is at the $\left(1 - \frac{1}{2}\alpha\right) M$ index of L.

3.3.2 Bradley-Terry Model

Bradley and Terry argued that the ratio's $\left(ij\right)$ satisfy a logistic density function instead of the normal density function proposed by Thurstone. They specify the ratio's $\left(ij\right)$ as follow

$$\begin{cases} (ij) = 1 - (ji)\,,\ 1 \le i,j \le n \\ \frac{r_i}{r_i + r_j} = (ij)\,,\ 1 \le i,j \le n \\ \sum\limits_{i=1}^{n} r_i = 1,\ 1 \le i \le n \end{cases} \tag{3.22}$$

where the r_i's are the underlying parameters of the logistic function and $r_i = V(i)$. The probability distribution used can be interpreted as a Bernoulli distribution, where the probability of success is given by $P(V(i) - V(j) > 0) = (ij)$ and probability of failure is given by $P(V(i) - V(j) < 0) = (ji)$.

Solution of the Bradley-Terry Model

For $n = 2$ the solution of the Bradley-Terry model is straight forward. It follows that $r_1 = (12)$ and $r_2 = (21)$. For $n \geq 3$ the system of equations becomes overdetermined and there are two possible ways to determine a solution. One is the method of least squares and the other is an iterative method, which can be found in the book of David [**?**].

The system of equation which seems non-linear in $V(i)$ and $V(j)$ can be rearranged, such that it becomes a system of linear equations, which is given as follows

$$\begin{cases} r_i\left((ij) - 1\right) + r_j\,(ij) = 0,\ 1 \le i,j \le n \\ \sum\limits_{i=1}^{n} r_i = 1,\ 1 \le i \le n \end{cases} \tag{3.23}$$

The least square solution of (**??**) $\hat{r}$, where $\hat{r}$ is a vector of estimated values of r_i, is given by $\hat{r} = \left(A^T A\right)^{-1} A^T b$, where

$$A = \begin{pmatrix} ((12)-1) & 12 & \cdots & \cdots & 0 \\ \vdots & & & & \vdots \\ \vdots & & ((ij)-1) & ij & \vdots \\ \vdots & & & ((n,n-1)-1) & (n,n-1) \\ 1 & \cdots & \cdots & \cdots & 1 \end{pmatrix} \text{ and}$$

$$b = \begin{pmatrix} 0 \\ \vdots \\ \vdots \\ \vdots \\ 1 \end{pmatrix}.$$

The iterative method uses the quantity π_i instead of the ratio (ij), where π_i is the number of times alternative i is preferred by some expert to some other alternative and k the number of individuals in the group. One seeks values for the r_i's such that under the modeling assumptions, the probability of seeing the ratio's (ij) is as large as possible[3]. From this David [?] formulates the maximum likelihood estimation of the parameters r_i as

$$r_i = \frac{\pi_i}{\sum_{j=1;j\neq i}^{n} k\left(r_i + r_j\right)^{-1}} \tag{3.24}$$

The solution of (**??**) can be found as follows. Begin with initial values for the r_i's e.g. $\frac{1}{n}$. While the stopping criteria is not met define the next iteration as follows

$$r_i^{next} = \frac{\pi_i}{\sum_{j=1;j\neq i}^{n} k\left(r_i + r_j^l\right)^{-1}} \tag{3.25}$$

where $l = 1$ if $j < i$ and $l = 0$ if $j > i$. The stopping criteria is defined as the norm between the differences of the next and previous values of the r_i's. It may be necessary to add an extra stopping criteria, (e.g. the number of iterations smaller than a specified number, to prevent staying in the while loop).

Goodness of Fit

The least square solution $\hat{r}$ is an estimated of the true values r. One way of looking how good the estimated solution $\hat{r}$ fits the true values of r is by looking at the adjusted coefficient of determination $R^2_{adjusted}$. The adjusted coefficient of determination $R^2_{adjusted}$ is used instead of the coefficient of determination $\left(R^2\right)$ [4], because the system of equations will have a degree of freedom higher than one for $n > 2$. When there are more than one degree of freedom the coefficient of determination would tend to come artificially high simply because some independents chance variations "explain" small parts of the variance of the dependent. A coefficient of determination close to zero would explain a bad fit of the data to the model, where as a coefficient of determination close to one would explain a good fit of the data to the model. The coefficient of determination is defined as, $R^2 = \left(1 - \left(\frac{SSE}{SST}\right)\right)$, where $SSE = error\ sum\ of\ squares = \sum_{i=1}^{(n-choose-2)+1} \left(\hat{Y}_i\right) - \bar{Y}^2$, Y_i is the actual value of Y for the ith case,

[3] This is known as the maximum likelihood solution. (Ford 1957[?]) has shown that the solution is unique and that the iterative process converges to this solution if the following condition is met: it is not possible to divide the set of alternatives into two nonempty subsets, such that no alternative in one subset is preferred by any respondent above some alternative in the second subset

[4] also called multiple correlation or the coefficient of multiple determination, is the percent of the variance in the dependent explained uniquely or jointly by the independents. R^2 can also be interpreted as the proportionate reduction in error in estimating the dependent when knowing the independents. That is, R^2 reflects the number of errors made when using the regression model to guess the value of the dependent, in ratio to the total errors made when using only the dependent's mean as the basis for estimating all cases.

$\hat{Y}_i$ is the regression prediction for the ith case, and where $SST = total\ sum\ of\ squares = \sum_{i=1}^{(n-choose-2)+1} \left(Y_i - \bar{Y}\right)^2$ with $\bar{Y} = \sum_{i=1}^{(n-choose-2)+1} Y_i$. The adjusted coefficient of determination is defined as

$$R^2_{adjusted} = 1 - \left(1 - R^2\right)\left(\frac{(n - choose - 2) - 1}{(n - choose - 2) - n - 1}\right) \tag{3.26}$$

where in this research $(n - choose - 2)$ is the number of equations and n the number of alternatives.

The goodness of fit for the iterative solution (**??**) is given as follows

$$F = 2\left(\sum_{\substack{i,j=1 \\ i \neq j}}^{n} \pi_{ij} \ln\left((ij)\right) - \sum_{i=1}^{n} \pi_i \ln\left(\hat{r}_i\right) + \sum_{i<j}^{n} n \ln\left(\hat{r}_i + \hat{r}_j\right)\right) \tag{3.27}$$

where π_{ij} is the number of times alternative i is preferred to alternative j by some expert. The quantity F is asymptotically chi square distributed with $\frac{(n-1)(n-2)}{2}$ degrees of freedom (Bradley[**?**]).

Confidence Bounds

The simulation method (bootstrap) for determining the confidence bounds of the values obtained by the Bradley-Terry model is given as follow

Init:
solve model with given data and obtain values for parameters of the model. Define a list L that will store parameter values, a list K to store the calculated ratio's $(ij)^{'}$, the number of runs M, and a significance level α .
For:
$k = 1..M$
Run:
Use the $\hat{r}_i$'s obtained from *Init* to define the ratio's $(ij)^{'}$. For each individual and each comparison draw a sample from the Bernoulli distribution with parameter $(ij)^{'}$ and calculate the next ratio's $(ij)^{next}$ and store them in K. Solve the model again with $(ij)^{next}$ and store $\hat{r}_i$'s in L.
Finalize:
Sort the values for each parameter. The confidence bound for each parameter is obtained by selecting the lower bound, which is at the $\frac{1}{2}\alpha M$ index of L and the upper bound, which is at the $\left(1 - \frac{1}{2}\alpha\right) M$ index of L.

3.4 Non-Parametric Models

The non-parametric models are inspired by a working paper of Roger Cooke called "Discrete Choice with Probabilistic Inversion". There is data available of the form:

$$Pr\,(i > j) = p_{ij}; i, j = 1...n; i \neq j; p_{ij} = 1 - p_{ji} \tag{3.28}$$

where p_{ij} is the probability that alternative i is preferred over alternative j.

A set of utility functions on a finite set is *standardized* if the functions have the same zero and unit. If the utility functions are standardized, then they take values in the interval $[0, 1]$. Hence the set U of utility functions becomes the set $\Omega = [0, 1]^n$. A utility function $u \in \Omega$ *expresses a preference* $\{i > j\}$ if $u\,(i) > u\,(j)$. A probability distribution Pr on Ω *expresses a set of pairwise preferences* if:

$$\forall i, j; Pr\,\{u \in \Omega | u\,(i) > u\,(j)\} = p_{ij} \tag{3.29}$$

Similarly, a probability distribution Pr over the set $n!$ of permutations of $\{1, ..., n\}$ *expresses a set of pairwise preferences* if

$$\forall i, j; Pr\,\{\pi \in n! | \pi\,(i) > \pi\,(j)\} = p_{ij} \tag{3.30}$$

where $\pi\,(i) > \pi\,(j)$ is the union of all the orderings where i is ranked higher than j. For n alternatives there are $\binom{n}{2}$ sets of the form $\pi\,(i) > \pi\,(j)$.

Problem 1
Find a probability measure on the set of orderings Ω that expresses the pairwise preferences $\{p_{ij}\}$. If there is more than one, or none, find the "best".

Problem 2
Find a probability measure on the set of utility functions Ω that expresses the pairwise preferences $\{p_{ij}\}$. If there is more than one, or none, find the "best".

Clearly, a solution to the first problem yields a solution to the second and conversely. In fact, if the probability for a given ranking $\pi \in n!$ is distributed evenly over all $u \in \Omega$ which express that ranking, then this will yield the minimum information distribution with respect to the Lebesgue measure which satisfies (**??**).

Solution for Problem 2

First there must be proved that the probability measure on the set Ω is indeed a probability measure. Given an event E in Ω which is either finite with N elements or countable infinite with ∞ elements, then one may write

$\Omega \equiv \bigcup_{i=1}^{N} E_i,$

and a quantity $Pr\,(E_i)$, called the probability of event E_i, is defined such that

1. $0 \leq Pr\left(E_i\right) \leq 1$
2. $P\left(\Omega\right) = 1$
3. *Additivity* $: Pr\left(E_i \bigcup E_j\right) = Pr\left(E_i\right) + Pr\left(E_j\right),$
   $\forall i, j$ *where* E_i *and* E_j *are mutual exclusive*
4. *Countable additivity* $: Pr\left(\bigcup_{i=1}^{N} E_j\right) = \sum_{i=1}^{N} P\left(E_i\right)$ *for* $n = 1, 2, ..., N,$
   where $E_1, E_2, ...$ *are mutual exclusive* $(i.e.E_1 \bigcap E_2 \equiv \emptyset)$

The outcome space Ω of (**??**) is a finite set with $n!$ elements. The events $\{\pi\left(i\right) > \pi\left(j\right)\}$ and $\{\pi\left(j\right) > \pi\left(i\right)\}$ are mutually exclusive and the probability of these events are marginal distributions. From this it clear that

1. $\forall \pi \in \Omega \ \pi \in \{\pi\left(i\right) > \pi\left(j\right)\}$ *or* $\pi \in \{\pi\left(j\right) > \pi\left(i\right)\}$
2. $Pr\left(\{\pi\left(i\right) > \pi\left(j\right)\} \bigcup \{\pi\left(j\right) > \pi\left(i\right)\}\right) = Pr\left(\{\pi\left(i\right) > \pi\left(j\right)\}\right) + Pr\left(\{\pi\left(j\right) > \pi\left(i\right)\}\right) = p_{ij} + 1 - p_{ij} = 1$

The only detail left for Pr to be a probability measure is $\forall \pi \in \Omega : \ 0 \leq Pr\left(\pi\right) \leq 1 \wedge Pr\left(\bigcup \pi\right) = 1.$

The problem formulated in (**??**) can be written as a linear problem with constraint $\forall \pi \in \Omega : \ 0 \leq Pr\left(\pi\right) \leq 1 \wedge Pr\left(\bigcup \pi\right) = 1.$

$$\begin{cases} max \sum_{\pi \in \Omega} Pr\left(\pi\right) \\ \sum_{\pi \in \{\pi(i) > \pi(j)\}} Pr\left(\pi\right) = p_{ij} \\ \sum_{\pi \in \{\pi(j) > \pi(i)\}} Pr\left(\pi\right) = p_{ji} = 1 - p_{ij} \\ \forall \pi \in \Omega : \ 0 \leq Pr\left(\pi\right) \leq 1 \end{cases} \tag{3.31}$$

If system (**??**) has a solution, which satisfies $\sum_{\pi \in \Omega} Pr\left(\pi\right) = 1$ then Pr is a probability measure over Ω. Now the set of orderings $\{\pi \in n!\}$ has a $one-to-one$ relation with the set of permutation matrices $P_\pi \in \{0,1\}^n \times \{0,1\}^n$. The solution can be written as follows

$$D = \sum_{k=1}^{n!} p_k P_{\pi_k} \tag{3.32}$$

where p_k is the probability for ordering k. The matrix D is a doubly stochastic[5], because each permutation matrix P_{π_k} is double stochastic and $\sum_{k=1}^{n!} p_k = 1$. Further the entry $i, j \in n$ correspondents with the probability that alternative i is ranked j out of n.

Matrix D can be associated with the transition matrix $\mathbf{P}$ of a Markov chain process with entries $P_{ij} = Pr\left(X_{n+1} = j | X_n = i\right)$ where P_{ij} is the probability

[5] A double stochastic matrix, is a matrix where the rows and the columns sum up to one

that alternative i goes to rank j in the next stage.

The stationary distribution ρ of the Markov chain process then yields a ranking over the alternatives. The following system of equation needs to be solved to find a stationary distribution ρ

$$\rho^T \mathbf{P} = \rho^T \tag{3.33}$$

The system of equation (**??**) need not have a solution for a general transition matrix $\mathbf{P}$ and if it does the solution does not have to be unique. However, if the transition matrix $\mathbf{P}$ is irreducible and aperiodic, then there exists a unique stationary distribution μ. In addition, $\mathbf{P}^k$ converges element wise to a rank-one matrix in which each row is the (transpose of the) stationary distribution ρ^T, that is

$$\lim_{k \to \infty} \mathbf{P}^k = \mathbf{1}\rho^T \tag{3.34}$$

where $\mathbf{1}$ is a $1 \times n$ matrix with all entries equal to one. If all entries of P_{ij} are positive then the transition matrix $\mathbf{P}$ is irreducible and aperiodic.

Solution for Problem 1

Finding a solution for the problem defined in (**??**) is by formulating the problem as a constrained optimization problem. The problem is defined as follows:

Let μ be the Lebesgue measure on $[0,1]^n$, find a probability density f on $[0,1]^n$ such that it satisfies the marginal distribution $\int_{\{u|u(i)>u(j)\}} f d\mu\,(x) = p_{ij}$. The problem is formulated as

$$\begin{cases} min \int_{[0,1]^n} ln\,(f\,(x))\, d\mu(x) \\ \int_{A_{ij}} f\,(x)\, d\mu(x);\, A_{ij} = \{u|u\,(i) > u\,(j)\}\,. \end{cases} \tag{3.35}$$

Determining the feasibility of (**??**) is as hard as solving it. Cooke[**?**] formulated **problem 1** as a probabilistic inversion problem.

Let $C:\ \Omega \to \{0,1\}^{n-choose-2}$ be defined as the mapping $C\,(u) = (C_{12}\,(u)\,, ..., C_{n-1,n}\,(u))$ and a measure $Pr = \prod Pr_{ij}$ on $\{0,1\}^{n-choose-2}$. The problem is to find a measure λ on Ω, minimally informative with respect to the Lesbesgue measure, such that $C\,(\lambda) = Pr$, where $C\,(\lambda)$ is the "push through" of λ onto $\{0,1\}^{n-choose-2}$. In other words

$$C\,(\lambda)\,(B) = \lambda\left(C^{-1}\,(B)\right) \forall \subseteq \{0,1\}^{n-choose-2} \tag{3.36}$$

Equation (**??**) can be solved using two algorithms based on sample re-weighting. A large number of samples is drawn from Ω and the samples are re-weighted in order to satisfy the constraints. The two algorithms that implement the re-weighting are known as IPF (Iterative Proportional Fitting) and

PARFUM(Parameter Fitting For Uncertain Models).

A result with one of the iterative solvers is obtained as follows. First create a set of n vectors where n is the number of alternatives. Draw a large number of samples M from n independent uniform distributions and store the samples of each uniform distribution in each of the n vectors. From this create $\frac{1}{2}n\,(n-1)$ indicator vectors $1_{\{u_i>u_j\}}$. These indicator vectors are the variables for either the IPF or PARFUM solvers. The quantiles are given by $P\left(\{u_i<u_j\}\right)$ and the quantile points are equal to zero. The purpose it to re-weight the samples from the variables with one of the iterative solvers such that the value "0" is equal to (ji). The convergence of the PARFUM solver is far slower than the IPF solver for this case. A stopping criteria for both solvers is to look at the L^1 norm of the difference $W_{k+1}-W_k$ where W is the joint distribution. The initial joint distribution is given by $\frac{1}{M}$. When the norm of the difference $W_{k+1}-W_k$ increases again for the IPF solver it means that the IPF solver has reached a solution, because after that the IPF solver will start oscillating.

The following can be said about the expected values of the sum of u_i's. Consider $u_1\cdots u_n$. Each permutation of $1\cdots n$ can be associated with a hypercube in $[0,1]^n$. For example $(2,3,7,1\cdots)$ correspondents with $u_2>u_3>u_7>u_1\cdots$. One can verify the following

$$\sum_{i=1}^{n}\int_0^1\int_0^{u_1}\int_0^{u_2}\cdots\int_0^{u_n}u_i du_n\cdots du_2du_1=\frac{\frac{n}{2}}{n!} \tag{3.37}$$

where $\frac{1}{n!}$ is the mass in a permutation hypercube if $u_1\cdots u_n$ are independent uniform on $[0,1]$. Now, IPF and PARFUM assign weights in a manner that only depends on the order of the variables. I.e. two samples in which the ui's have the same order get the same weight. Conditional on the order hypercube, the expectation of the sum of the u_i's is thus $\frac{n}{2}$, and summing over all hypercubes, the sum expectation is $\frac{n}{2}$.

3.5 Finding a Set of Explanatory Variables

Aggregated subjective scores for countries based on definition global competitiveness can be obtained using one of the previously described models. Once these scores are acquired one might search for a set of data that best reflects these scores. For the study the set of data will consist out of the nine pillars of the World Economic Forum (WEF). Regression is a useful tool for determining how well a set of data reflects these scores.

In regression the scores are referred to as the dependent variable and the set of data is referred to as the independent or explanatory variables. From this the following relation between the dependent variable and independent variables can be formulated

$$\mathbf{y} = F(\mathbf{x}) \tag{3.38}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \ \mathbf{x} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1m} \\ \vdots & \vdots & & \vdots \\ 1 & x_{1m} & \dots & x_{nm} \end{bmatrix}, \ F : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$$

The simplest mapping will be used and that is a linear mapping, which leads to linear regression

$$\mathbf{y} = \mathbf{x}\mathbf{b} + \epsilon \tag{3.39}$$

where

$$\mathbf{b} = \begin{bmatrix} b_0 \\ \vdots \\ b_m \end{bmatrix}, \ \epsilon = \begin{bmatrix} \epsilon_0 \\ \vdots \\ \epsilon_m \end{bmatrix}$$

The b's are the weights and also known as regression coefficients and the ϵ's the errors made by estimating the relationship between $\mathbf{y}$ and $\mathbf{x}$ with a linear function. It is assumed that the errors are normally distributed with mean zero and variance $\frac{\sigma^2}{k}$. The goal is to find a set of weights $\mathbf{b}$, which minimize the error. The weights $\mathbf{b}$ using least squares:

$$\mathbf{b} = \left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{y} \tag{3.40}$$

The matrix $\left(\mathbf{x}^T\mathbf{x}\right)$ will always be invertible if the columns of $\mathbf{x}$ are linear independent. In the case of the research this will always hold if the number of explanatory variables is smaller than the number of countries.

The coefficient of determination R^2 determines how good the data fits the model and is given by

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \tag{3.41}$$

where $SSE = \textit{error sum of squares} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$, y_i is the actual value of $\mathbf{y}$ for the ith case, $\hat{y_i}$ is the regression prediction for the ith case, $SST = \textit{total sum of squares} = \sum_{i=1}^{n} (y_i - \bar{y})^2$ with $\bar{y} = \sum_{i=1}^{n} y_i$, and $SSR = \textit{sum of squared residuals} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$.

The adjusted coefficient of determination $R^2_{adjusted}$ gives a better picture how well the data fits the model when the number of degrees of freedom becomes large. The formula of $R^2_{adjusted}$ is the same as in equation (??). The set of explanatory that best describe the scoring is a permutation of variables with the highest $R - squared$ value. For m explanatory variables there will be $m!$ of these permutations, but the test can easily be programmed and run on a computer.

3.5.1 Correlation among explanatory variables

The objective is to find a set of **independent** explanatory variables that **best** reflect the scoring. Previously was showed that $R - squared$ gives a measure of how good the explanatory variables linearly describe the ranking. However it was not shown how to determine if the explanatory variables are independent. The explanatory variables will always have some kind of dependency. If explanatory variable x_{j+1} is highly correlated with x_j then one of them can be removed from the set of explanatory variables that **best** describes the ranking.

The **correlation coefficient** $r^2_{x_j x_{j+1}}$ measures the dependence between the two variables x_j and x_{j+1}, and is given by

$$r^2_{x_j x_{j+1}} = \frac{\left(\sum\limits_{i=1}^{n} (x_{ij} - \bar{x}_j)(x_{i,j+1} - \bar{x}_{j+1})\right)^2}{\sum\limits_{i=1}^{n} (x_{ij} - \bar{x}_j)^2 \sum\limits_{i=1}^{n} (x_{i,j+1} - \bar{x}_{j+1})^2} \tag{3.42}$$

If $r^2_{x_j x_{j+1}}$ is close to one then the variables x_j and x_{j+1} are highly correlated and a value close to zero means that x_j and x_{j+1} are poorly correlated. For high values of $r^2_{x_j x_{j+1}}$ the analyst should consider dropping either x_j or x_{j+1}.

3.5.2 Confidence Bounds Regression Coefficients

Confidence Bounds of the regression coefficient from the linear model (??) by using the **central limit theorem**. The **central limit theorem** states that given a distribution X with mean μ and variance σ^2, the **sample distribution** of the mean approaches a *normal distribution* with mean $\hat{\mu}$ and variance $\frac{\sigma^2}{n}$ as n, the sample size increases. The central limit theorem can be formulates as follows

$$\lim_{n\to\infty} Pr\left(\frac{\frac{1}{n}\sum\limits_{i=1}^{n} X_i - \mu}{\sigma/\sqrt{n}} \le z\right) = \Phi(z) \tag{3.43}$$

where Φ is the standard normal function. Wooldridge[?] shows that under the **central limit theorem** the regression coefficients are normally distributed with mean b_j and $Var\left(\hat{b}_j\right)$ for $j = 1..m$, with

$$Var\left(\hat{b}_j\right) = \frac{\sigma^2}{SST_j\left(1 - R_j^2\right)} \tag{3.44}$$

, where $SST_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$ is the total variation in x_j, and R_j^2 is the $R-squared$ from regressing x_j on all the other explanatory variables. The term σ^2 is the expected value of the total variance of the error terms ϵ. Unfortunately σ^2 can not be used, because the error terms are not observed. What can be observed is $\hat{\epsilon}_i = (y_i - \hat{y}_i)$ for $i = 1..n$, so σ^2 is replaced by $\hat{\sigma}^2$, with $\hat{\sigma}^2 = \left(\sum_{i=1}^{n} \hat{\sigma}_i^2\right)/(n - m - 1) = SSR/(n - m - 1)$. The term $n - m - 1$ is the **degrees of freedom** (*df*) for the linear regression with n observations and m explanatory variables. The *SSR* is divided by *df*, because otherwise $\hat{\sigma}^2$ is influenced by the number of observations n and number of explanatory variables m.

4. DATA ACQUISITION

The research requires data to validate the models. The process of the data acquisition and manipulation is organized with the aid of specific software tools. Three stages can be identified in the process of the data acquisition and manipulation. In stage one data is acquired from the individuals and stored into a database.

The second stage is the stage where the data acquired from the individuals should be analyzed for significant difference in preference for each individual and for significant difference in preference for the individuals as a group. After these analyses the data should be organized, so that it can be used as an input for each of the models. Initially each of these models only need the ratio's (ij), but the iterative solver of the Bradley-Terry model also needs π_i (the number of times alternative i is preferred to the other alternatives).

Finally the results of each model, which is again data should be presented, so that it is clear and apprehensible. The reason why the results need to be clear and comprehensible is that the top-down approach formulated in this research should be transparent and there should be room for discussion.

4.1 Data Stage One

In this research the concept global competitiveness is investigated, so the objective is to find preferences of individuals around the world. The data of stage should consist of the preferences of these individuals willing to participate in the research. It is also necessary to keep track of the individuals, because some of these individuals might be asked to participate in future studies. The internet makes it possible to reach individuals globally, whenever these individuals have access to an internet connection. One of the technologies of the internet that can make this all possible is a website. It is also possible to reach individuals via e-mail, but one of the objectives was to keep the effort of the analysts for obtaining the data the same independent of the number of individuals. E-mail would require an analyst to periodically check and group data, which will increase the effort if the number of periods increases. Also an individual should be able to carry out part of the comparisons and carry out the rest of the comparisons at a later time, which again advocates the use of a website. At last the data of stage three should be presented to invoke discussion and a website is again the best solution for doing this.

The website is required to be dynamic, because it should acquire and present data. However the data acquired from stage one should be stored in order to be used in stage two. In this research the data is stored in a database server MySQL officially pronounced as "My Ess Queue Ell" and handled by PHP[1]. The data to be stored for each individual are the individual's information, e.g. surname, firstname, background, country, and email address, and the comparisons they carry out. Therefore two MySQL tables are created to store the individual's details and the comparisons carried out by the individual.

For the research some extra tables where created to easily maintain a study or project. There is a table created that stores the alternatives and a table that stores the project details. The following scheme shows the tables created and the relations between the tables. Note that the tables expert, alternative, and comparison all have wrapper tables experts, alternatives and comparisons this is done to better visualize the tables and there relations. In the database these wrappers are not present.

4.2 Data Stage Two

In stage two data is required for the input of the models. Each model need data from the alternatives, the individuals (in the scheme revert to as experts) and the comparisons. Each data table from stage one has an identifier (ID), which makes the data easy to work with. Each individual who is willing to carry out comparisons for a project has provided entries in the comparison data table of stage one. The result of each comparison can have three different values, the default value zero (when the comparison is not carried out), one if the alternative i is preferred to alternative j and minus one alternative i is not preferred to alternative j. The comparisons of each of the individuals are transformed to a matrix with size the number of alternatives in the project. If the result of comparison i, j is one then entry i, j is one and if the result is minus one then the entry j, i is one. The frequency matrix that can be used by each model is the sum of all the matrices of the individuals.

4.3 Data Stage Three

Results are obtained from the models and these result need to be presented. The output of each of the models is saved to an XML[2] file. In return the xml files

[1] PHP recursive acronym for **H**ypertext **P**reprocessor is a widely-used general-purpose scripting language that is especially suited for Web development and can be embedded into HTML

[2] Extensible Markup Language (XML) is a W3C-recommended general-purpose markup language for creating special-purpose markup languages, capable of describing many different kinds of data. In other words: XML is a way of describing data and an XML file can contain the data too, as in a database

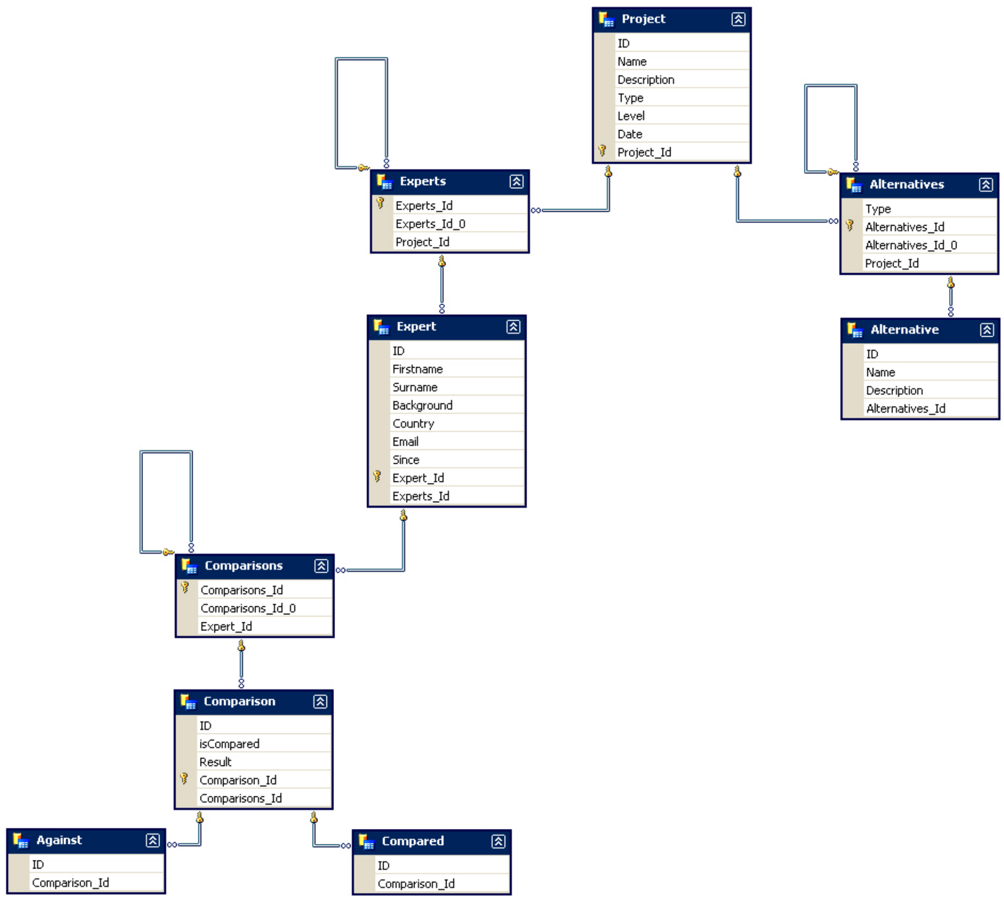

Fig. 4.1: The structure and relations of the tables

are stored in the database. These files can be accessed and displayed any time on the website. XSL (eXtensible Stylesheet Language) is used for displaying the XML files, which is a family of languages that allows one to describe how files encoded in the XML standard are to be formatted or transformed.

5. IMPLEMENTATION OF THE RESEARCH

For the three stages of the research, a website, a database server, and a program are used. The database server functions as an intermediate between the website and the program and is responsible that all the data gets handled. This can be visualized by the following scheme.

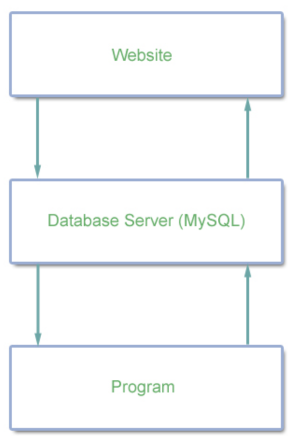

Fig. 5.1: The process of the three stages

Taking a look at figure (**??**) will suggest that there are users operating on the website and users operating on the program. The tasks and purposes of these users are distinct and the website and program help keeping the distinction. The users operating on the website are the individuals that are willing to participate (providing for data of a project) in a project or that seek information. The users operating on the program are the ones that want to conduct a study or project (gathering and analyzing the data).

To establish communication between the website and the database server **PHP** is used and the communication between the database and the program is done by **C-sharp**. Technically the communication to the database is done by **SQL**(Structured Query Language), but for example **PHP** is responsible to translate information or data from the website to **SQL** and from **SQL** back to information requested by the website. In theory there are several programming languages that are able to do the jobs described. **PHP** is chosen to develop the website, because it is a well documented scripting language, it is open source, enables the use of classes, finding errors is easier than for example **ASP**(Active Server Page), and last but not least **PHP** is supported on several operating systems. The last point was one of the crucial point for choosing **PHP**, because the webservers of the **Joint Research Center** (where the research is conducted) run under the operating system **Linux**. Choosing for example **ASP.NET** would require to install **IIS** webserver of **Microsoft**. The almost the same preferences were taking in to account when **MySQL** is selected as the database server. However for programming the program **C-sharp** was chosen, which is a programming languages that mainly runs on the operating systems of **Microsoft**. **Microsoft** has recently developed a framework **.NET**, which makes the programming under **Windows**(operating systems of **Microsoft**)

5.1 The Website

The choice was made for a website, because a website is by definition accessible by anyone with a internet connection. The more respondents carry out the comparisons the better the data will fit the models. Reaching these respondents is simplified by the use of the website. The website has been structured in three specific section. The section where general information can be found about what ranking methods are and how they are used. This section is placed as the homepage of the website. This section also contains a subsection news which informs the visitors about ongoing projects. The website can be visited by going to the following link `http://subjective-measures.jrc.it`

5.2 The program

There are software programs that can analyze paired comparison data using either the Bradley-Terry or Thurstone model for example **Excalibur**, but in this research the choice was made to develop a new software program. The reason was that the paired comparison data was collected and stored in a database and the program should be able to capture the data from the database. The current version of **Excalibur** does not have support for capturing data from a database nor from for example **XML**(eXtendible Markup Language).

The program used in this research is written from scratch using the theory behind this research. The programming paradigm used to develop the program is **OOP** (object-oriented programming), which is widely used in current software development. The idea behind object-oriented programming is that a computer program may be seen as comprising a collection of individual *units*, or *objects*, that act on each other, as opposed to a traditional view in which a program may be seen as a collection of functions, or simply as a list of instructions to the computer. The **OOP** approach is often simpler to develop and to maintain, lending itself to more direct analysis, coding, and understanding of complex situations and procedures than other programming methods. **OOP** has the following concepts

- *Class* - the unit of definition of data and behavior (functionality) for some kind-of-thing. For example, the 'class of Dogs' might be a set which includes the various breeds of dogs. A class is the basis of modularity and structure in an object-oriented computer program. A class should typically be recognizable to a non-programmer familiar with the problem domain, and the code for a class should be (relatively) self-contained and independent (as should the code for any good pre-OOP function). With such modularity, the structure of a program will correspond to the aspects of the problem that the program is intended to solve. This simplifies the mapping to and from the problem and program.

- *Object* - an instance of a class, an object (for example, "Lassie" the Dog) is the run-time manifestation (instantiation) of a particular exemplar of a class. (For the class of dogs which contains breed types, an acceptable exemplar would only be the subclass 'collie'; "Lassie" would then be an object in that subclass.) Each object has its own data, though the code within a class (or a subclass or an object) may be shared for economy. .

- *Method* (also known as message) - how code can use an object of some class. A method is a form of subroutine operating on a single object. Methods may be divided into queries returning the current state and commands changing it: a Dog could have a query Age to say how old it is, and command chase (Rabbit target) to start it chasing a rabbit. A method may also do both, but some authorities (e.g. Bertrand Meyer) recommend they be kept separate. Sometimes access to the data of an object is restricted to the methods of its class

- *Inheritance* - a mechanism for creating subclasses, inheritance provides a way to define a (sub)class as a specialization or subtype or extension of a more general class: Dog is a subclass of Canidae, and Collie is a subclass of the (sub)class Dog. A subclass inherits all the members of its superclass(es), but it can extend their behavior and add new members. Inheritance is the "is-a" relationship: a Dog is a Canidae. This is in contrast to composition, the "has-a" relationship: a Dog has a mother (another Dog) and has a father, etc

- *Encapsulation* - ensuring that code outside a class sees only functional details of that class, but not implementation details. The latter are liable to change, and could allow a user to put an object in an inappropriate state. Encapsulation is achieved by specifying which classes may use the members of an object. The result is that each object exposes to any class a certain interface - those members accessible to that class. For example, an interface can ensure that puppies can only be added to an object of the class Dog by code in that class. Members are often specified as public, protected and private, determining whether they are available to all classes, sub-classes or only the defining class. Some languages go further: Java uses the protected keyword to restrict access also to classes in the same package, C-sharp and VB.NET reserve some members to classes in the same assembly using keywords internal (C-sharp) or Friend (VB.NET), and Eiffel allows one to specify which classes may access any member.

- *Abstraction* - the ability of a program to ignore the details of an object's (sub)class and work at a more generic level when appropriate; For example, "Lassie" the Dog may be treated as a Dog much of the time, but when appropriate she is abstracted to the level of Canidae (superclass of Dog) or Carnivora (superclass of Canidae), and so on.

- *Polymorphism* - polymorphism is behavior that varies depending on the class in which the behavior is invoked, that is, two or more classes can react differently to the same message. For example, if Dog is commanded to speak this may elicit a Bark; if Pig is commanded to speak this may elicit a Grunt.

Each data tables from the data scheme (**??**) presented in the previous chapter can be translated to classes using **OOP** and will form the core of the new program. Note that not all the classes of the program are included in the data scheme (**??**) for example the program will also contain a class of solvers, a class of graphical user interfaces (GUI) and many other classes.

6. RESULTS

The **Joint Research Center of the European Commission** where the study was carried out has several Units consisting out of team of researchers. The researchers of the Unit **Econometrics and Statistical Support to Anti-Fraud** were asked to participate in the study. The other part of participants came from outside the Unit and in total 36 people participated in the study. The number of alternatives, which in the case of this study are countries, is eight.

The countries are selected from the three mayor regions of the world namely America, Asia and Europe. Each of the countries is given an ID this is merely done to simplify the programming and the storing of the countries. The ID's are numerical and must not be confused with the ranking of the countries. The countries are given by the following table.

ID	Country	Region
1	Finland	Europe
2	United States of America	America
3	United Kingdom	Europe
4	Italy	Europe
5	Hungary	Europe
6	China	Asia
7	Mexico	America
8	South-Korea	Asia

Tab. 6.1: The countries used in the study

For obtaining a scoring over the eight countries 28 comparisons need to be carried out by the participants. After each participant carried out all of the 28 comparisons they were able to give there comment about the study. The following are some of the participants' comments about the project.

> *"I think that the question should be segmented by sector, and by capital or labour intensive industries. If not, one can not reply. Ciao!"*

> *"The point of the pairwise comparisons is not clear. What is the list of criteria upon which to base the decision? For which sectors? Maybe a list of the indicators included in the GCI would be useful."*

> *"I looking forward to the results. Very interesting. I was wondering if there is any other background information into which methods are being used during the project"*

> *"Thanks! Was not so hard for an Economist reader. Good luck with your thesis!"*

These comments are interpreted as the extremes of all the comments, where no comments are seen as positive feedback. To the first two comments presented one might answer back that the whole idea behind the research was to obtain an index, which is an aggregation of several aspects. On the other hand one should also admit that for none experts it might be useful to provide information about the aspects, which might be covered by the index.

After analyzing the comments the program can be executed. The program starts with given a *treeview* of the study, which will be referred to as a project. When clicking on an item of the *treeview* detailed information is given about the item. The *treeview* is shown in the following picture.

6.1 Result Preliminary Analysis

The theory behind this research mentioned to look for the significant difference in preference both for each individual (participant) as for the group. The number of triads for each individual can be viewed by selecting an individual in the treeview. It is up to the analyst to decide at which level α to reject an individual which has a number of triads that exceed for the given α. In this study $\alpha = 0.01$ and for eight alternative the maximum number of triads should not exceed four, which resulted that the preferences of 27 individuals passed the analysis. The analyst has the possibility to use a different α and or to exclude some respondents from the study. Not all participating respondents completed all of the comparisons and for that reason are excluded from the study. Participants that have not completed all of the comparisons can be detected by having a negative number of items.

The following tables show the number of triads and the number of individuals

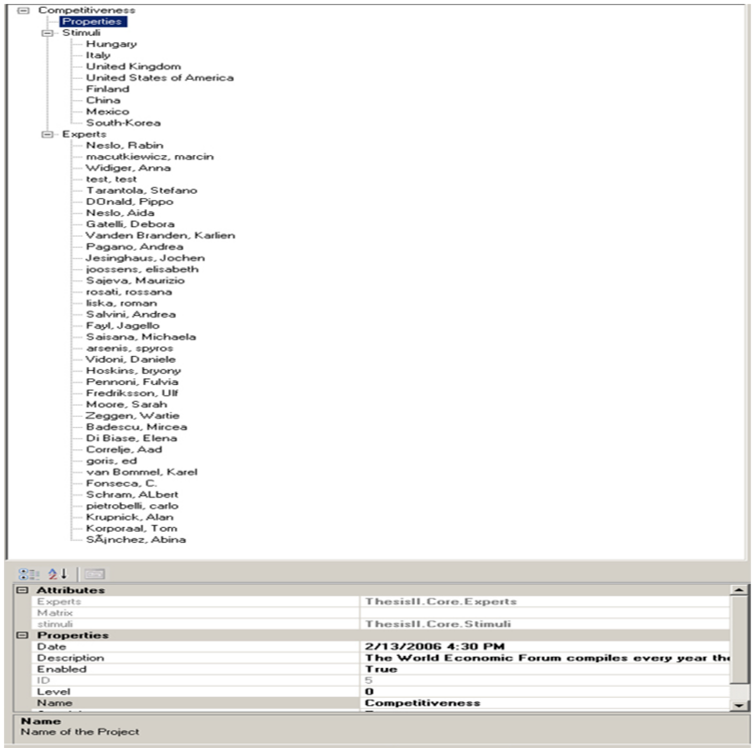

Fig. 6.1: Screenshot of Study

related.

Number of Triads	Number of Experts	p	$p^{'}$
0	19	1.000	0.99938
1	3	0.99985	0.99887
2	3	0.99955	0.99799
3	1	0.99987	0.99646
4	1	0.99972	0.99388
5	0	0.99936	0.98961
6	0	0.989	0.98269
7	1	0.977	0.97177
8	0	0.963	0.95499
9	0	0.937	0.92999
10	1	0.906	0.89400

Tab. 6.2: Number of Triads vs. Number of Experts

where p is the probability from table B.4 of Cooke[**?**] and $p^{'}$ is from using the fact that the quantity (**??**) defined in the theory is approximately distributed with $v = \frac{n(n-1)(n-2)}{(n-4)^2}$ degrees of freedom. The probabilities p and $p^{'}$ differ from each other, which results in the rejection of the preferences of an individual at another number of triads. Using the probability p would lead to rejection of the preferences of an individual with more than six triads instead of four.

The next analysis to be carried out is the significant difference for the group. The theory mentioned two test for this analysis namely the coefficient of agreement ϑ and the coefficient of concordance ω. The value of the coefficient of agreement ϑ for the 27 individuals of this study is $\vartheta = 0.163614163614164$. And the value of the coefficient of concordance ω is $\omega = 0.250404174015285$. Both values of the coefficients indicate that the preferences of the individuals do not agree.

The value for $\omega^{'}$ is equal to 47.3264 with eight countries the probability of seeing a value greater than $\omega^{'}$ is given by 0.99999, so the hypothesis that that the preferences are at random is rejected.

For $\vartheta^{'}$ of (**??**) a value of 160.0896 is obtained with 31.4496 degrees of freedom. The probability of seeing a value larger than $\vartheta^{'}$ is 1 , so the hypothesis that all agreements are due to chance is rejected.

6.2 Results Parametric Models

All the models depend on the preference matrix F where the entry F_{ij} means the number of times i is preferred to j. A ranking over the countries can be obtained by simply summing for each row over the columns of the matrix F.

Country	1	2	3	4	5	6	7	8	Score	Ranking
1	0	17	15	23	16	14	22	18	125	1
2	10	0	13	22	15	12	24	16	112	4
3	12	14	0	25	18	14	23	19	125	1
4	4	5	2	0	14	5	19	11	60	7
5	11	12	9	13	0	8	18	14	85	5
6	13	15	13	22	19	0	23	20	125	1
7	5	3	4	8	9	4	0	8	41	8
8	9	11	8	16	13	7	19	0	83	6

Tab. 6.3: Frequency matrix

The ranking obtained by just summing up the columns for each row is not a ranking on interval or ratio scale. This method is known as the Condorcet method.

6.2.1 Thurstone Model

From the frequency matrix $\mathbf{F}$ the matrices $\mathbf{P}$ and $\mathbf{X}$ can be obtained. The entries P_{ij} of the matrix $\mathbf{P}$ are the ratios that i is preferred to j. The entries X_{ij} of the matrix $\mathbf{X}$ are the values of the inverse normal cumulative distribution and the matrix $\mathbf{X}$ forms a basis for both the Thurstone cases.

A significance level of $\alpha = 0.01$ is taken to compute the confidence bounds of the scale values. Theoretically if one choose a significance level of α then one need to simulate with more then $M = \frac{2}{\alpha}$ runs. For the study a number of one million run is taken to simulate the confidence bounds. Note that it might be possible that an entry of the matrix simulated $\mathbf{P}'$ maybe equal to one for $i \neq j$. When this occurs one might choose to disregard this simulation or choose a method that can deal with such kind of anomalies. In this study simulation with such anomalies are just left out. And there maybe people that argue that what can happen will happen, but for 27 individuals the chance that an entry can become equal to one is $\frac{1}{2}^{27} = 7.45058E-9$.

Running the program gives the following result for the first Thurstone model.

ID	Score	σ	Bounds score	α	Runs
3	0.60271725	1	[0.14928706 , 1.11417457]	0.01	1.0E6
6	0.55998279	1	[0.10205837 , 1.06762809]	0.01	1.0E6
1	0.55192401	1	[0.09528789 ,1.05298228]	0.01	1.0E6
2	0.34883894	1	[-0.10978906 ,0.84860887]	0.01	1.0E6
5	-0.16069219	1	[-0.65367735 ,0.30984858]	0.01	1.0E6
8	-0.19848377	1	[-0.69435714 ,0.26466735]	0.01	1.0E6
4	-0.69557611	1	[-1.23030618 ,-0.24251871]	0.01	1.0E6
7	-1.00871093	1	[-1.53543638 ,-0.55255648]	0.01	1.0E6

Tab. 6.4: Results first Thurstone model

It is advised to do the goodness of fit test before determining the confidence bounds of the scores, because it may occur that the model is not a good fit. Mosteller created a goodness of fit test for the Thurstone C case. He defined the quantity D (**??**) and argued that this quantity D is approximately chi-squared distributed with $\begin{pmatrix} n \\ 2 \end{pmatrix} - n + 1$ degrees of freedom. Further he argued that if the probability of seeing D is smaller than 0.05 then the first Thurstone model does not fit the data well.

The quantity for data of this study is $D = 14.64438$ with 21 degrees of freedom. The probability for seeing this quantity $D \geq 14.64438$ is equal to $p = 0.15964$. This probability is higher than the 0.05 so one might safely conclude that the first Thurstone model is a good fit to the data of this study.

Excalibur gives a different quantity $D = 13.35791$ this, because Excalibur uses a *arctan* instead of *arcsin*.

The Thurstone case B is the Thurstone model with small difference in the deviation and constant correlation coefficient, where the correlation coefficient in this study is equal to zero. The results for the Thurstone case B are given by the following tables.

ID	Score	σ
1	0.824909832	1.268921541
6	0.798592963	0.98580097
3	0.611102281	0.556017339
2	0.36154684	0.63650582
5	-0.294812882	1.91755193
8	-0.29515126	1.183933855
4	-0.541018546	0.298416567
7	-1.465169229	1.152851975

Tab. 6.5: Results Thurstone case B

ID	Bounds score	Bounds σ	α	Runs
1	[0.31692530 ,1.96835227]	[0.616786615 ,2.23272070]	0.01	1.0E6
6	[0.39383916 ,1.79670824]	[0.454677311 ,1.74339408]	0.01	1.0E6
3	[0.29502117 ,1.27237567]	[0.232018583 ,1.03714535]	0.01	1.0E6
2	[0.01468227 ,0.93070412]	[0.226580586 ,1.12279839]	0.01	1.0E6
5	[-1.51201777 ,0.74941859]	[1.03582764 ,3.54678564]	0.01	1.0E6
8	[-1.16359234 ,0.30707523]	[0.55482140 ,2.040466292]	0.01	1.0E6
4	[-1.12426249 ,-0.36245861]	[0.090727539 ,0.673630391]	0.01	1.0E6
7	[-2.60876454 ,-1.00943508]	[0.543551819 ,1.679601318]	0.01	1.0E6

Tab. 6.6: Bounds Thurstone case B

6.2.2 Bradley-Terry Model

The Bradley-Terry model is the second parametric model and as mentioned before it also uses the frequency matrix $\mathbf{F}$ to obtain the scores for the countries. The theory behind this study formulated two solution strategies for the Bradley-Terry model. The iterative solution and the least square solution.

The iterative solution uses a vector π out of the frequency matrix F where the entry π_i is the total number of times that country i is preferred to all other countries. The least square solution uses the matrix $\mathbf{P}$ where the entries P_{ij} are the ratios that i is preferred to j. The results for the iterative solution of the Bradley-Terry model as for least square solution of the Bradley-Terry model are given by the following tables.

ID	Score	Bounds score	α	Runs
3	0.19894114	[0.11326151, 0.32983719]	0.01	1.0E6
1	0.19894113	[0.11369775, 0.32972564]	0.01	1.0E6
6	0.19894113	[0.11385430, 0.33123488]	0.01	1.0E6
2	0.15087104	[0.08132308, 0.2565627]	0.01	1.0E6
5	0.08666955	[0.041761430, 0.15743320]	0.01	1.0E6
8	0.08315920	[0.039447530, 0.15123081]	0.01	1.0E6
4	0.05066744	[0.020628995, 0.09762578]	0.01	1.0E6
7	0.03180939	[0.01057216, 0.06591742]	0.01	1.0E6

Tab. 6.7: Results iterative solution of Bradley-Terry model

ID	Score	Bounds score	α	Runs
1	0.20744344	[0.11727059, 0.34161081]	0.01	1.0E6
6	0.19679057	[0.10787708, 0.32522436]	0.01	1.0E6
3	0.18792482	[0.10251706, 0.31377009]	0.01	1.0E6
2	0.14620725	[0.07526662, 0.25254897]	0.01	1.0E6
5	0.09616650	[0.04391545, 0.17558777]	0.01	1.0E6
8	0.08624411	[0.03871788, 0.15922902]	0.01	1.0E6
4	0.04106514	[0.01389969, 0.08511379]	0.01	1.0E6
7	0.03377519	[0.01024880, 0.07257935]	0.01	1.0E6

Tab. 6.8: Results least square solution of Bradley-Terry model

The goodness of fit test for the iterative solution of the Bradley-Terry model determines the quantity $F = 2\left(\sum_{\substack{i,j=1\\ i\neq j}}^{n} \pi_{ij}\ln((ij)) - \sum_{i=1}^{n} \pi_i \ln(\hat{r}_i) + \sum_{i<j}^{n} n\ln(\hat{r}_i + \hat{r}_j)\right)$ where F is asymptotically chi square distributed with $\frac{(n-1)(n-2)}{2}$ degrees of freedom (Bradley[**?**]). For this study the following value for the quantity is determined $F = 14.15875$. And with 21 degrees of freedom the probability of seeing the quantity $F \geq 14.15875$ is equal to $p = 0.13732$, so it seems that the iterative approach of the Bradley-Terry model fits the data well.

The goodness of fit test of the least square solution of the Bradley-Terry model is defined by determining the quantity $R^2_{adjusted}$. The value for the quantity $R^2_{adjusted}$ is equal to $R^2_{adjusted} = 0.98436800711355776$, so one might safely conclude that the model is a good fit to the data.

6.3 Results Non-Parametric Models

There were two non-parametric models proposed to obtain a scoring over a set of alternatives. The first model was defined as a maximization problem. Unfortunately there are not enough resources to solve this model for the eight countries selected for the study[1]

6.3.1 Results Iterative Models

There are two iterative solvers for the non-parametric model namely the IPF solver and the PARFUM solver. If the problem is feasible the IPF is preferred to the PARFUM solver and in the case of infeasibility the PARFUM solver is preferred, because it will always converges (see paper Cooke[**?**]).

After feeding the solvers with the variables (the samples) the quantiles and quantile the following results are obtained for the IPF and PARFUM solver.

ID	Score	σ
1	0.625198	0.270081
3	0.625022	0.246108
6	0.623923	0.258498
2	0.573246	0.258274
5	0.461565	0.303379
8	0.452682	0.276947
4	0.358079	0.225431
7	0.281076	0.233941

Tab. 6.9: Results IPF model

[1] All the models are solved on a PC (Personal Computer) and to allocate for example a number as a double precision in the memory of the PC cost eight bytes of memory. For the optimization algorithm, a matrix with dimensions $(2n!+1)\times n!$ is needed. With eight countries the amount of memory needed to load this matrix is 24.23 Giga bytes of memory. Unfortunately today's PC only have an average of 0.5 Giga bytes of RAM, which maybe extended by the use of virtual memory, but most commercial compilers will throw an out of memory exception. It may be concluded that for more than seven alternatives the maximization problem becomes computational intractable.

ID	Score	σ
6	0.626513	0.257125
1	0.626253	0.269102
3	0.624483	0.246230
2	0.570792	0.259185
5	0.459479	0.302792
8	0.451996	0.276985
4	0.356015	0.224739
7	0.278915	0.234326

Tab. 6.10: Results PARFUM model

ID	1	2	3	4	5	6	7	8
1	1.00	0.05	0.04	0.13	-0.06	0.01	-0.02	0.01
2	0.05	1.00	0.05	0.14	-0.06	0.01	0.18	-0.01
3	0.04	0.05	1.00	0.32	-0.04	0.04	0.03	0.03
4	0.13	0.14	0.32	1.00	-0.08	0.09	0.08	0.01
5	-0.06	-0.06	-0.04	-0.08	1.00	0.05	-0.01	0.03
6	0.01	0.01	0.04	0.09	0.05	1.00	0.05	0.07
7	-0.02	0.18	0.03	0.08	-0.01	0.05	1.00	0.03
8	0.01	-0.01	0.03	0.01	0.03	0.07	0.03	1.00

Tab. 6.11: Product moment correlation matrix of IPF results

The number of samples for both solvers is $M = 200000$. The IPF solver reaches a solution at 112 iterations with an error of $1.4188855851564315E-18$. The PARFUM solver converges very slowly and the maximum number of iterations is set at 300 with an error of $1.3012582535956193E-09$. The error at iteration 299 for the PARFUM solver is $1.314704502123372E-09$, which gives a convergence rate of 1.010333. The following plot shows the convergence rate of the IPF solver versus the PARFUM solver.

ID	1	2	3	4	5	6	7	8
1	1.00	0.05	0.04	0.12	-0.06	-0.02	-0.02	-0.01
2	0.05	1.00	0.05	0.14	-0.05	-0.01	0.18	-0.02
3	0.04	0.05	1.00	0.319	-0.04	0.03	0.03	0.02
4	0.12	0.14	0.319	1.00	-0.08	0.07	0.07	0.01
5	-0.06	-0.05	-0.04	-0.08	1.00	0.04	-0.01	0.00
6	-0.02	-0.01	0.03	0.07	0.04	1.00	0.04	0.08
7	-0.02	0.18	0.03	0.07	-0.01	0.04	1.00	0.01
8	-0.01	-0.02	0.02	0.01	0.00	0.08	0.01	1.00

Tab. 6.12: Product moment correlation matrix of PARFUM results

6.4 Discussion of The Models

For non-parametric models it is not necessary to compute confidence bounds, because everything is known about the probability distributions of the scores. The iterative non-parametric models suggest that there is some correlation between the countries. A correlation coefficient in absolute value less than 0.20 indicates slight, almost negligible relationship. And a correlation coefficient in absolute value between 0.20 and 0.40 indicates definite but small relationship. The parametric models are formulated on the assumption that there is no correlation between the countries, which would an acceptable assumption if the correlation coefficient between the United Kingdom and Italy was less than 0.20.

Summing over the columns of the frequency matrix F acquired from the comparisons of the respondents indicated that there is no ranking difference between the countries China, Finland and the United Kingdom. The original Bradley-Terry model is the only model, which reflects the ranking acquired from summing over the columns. Both the iterative non-parametric models will produce a different top three of countries and this, because each time the random generator of a computer will produce a different set of samples.

As a closing remark one must notice that the results can easily vary by including or excluding a small subset of the respondents. Two respondents participated later with the study and without both of their comparisons the results for all the models show a clear distinction in ranking over the countries. The only results that stay stable in ranking are the results from the Thurstone B case (ThurB) and the least square solution of the Bradley-Terry model (B-TII).

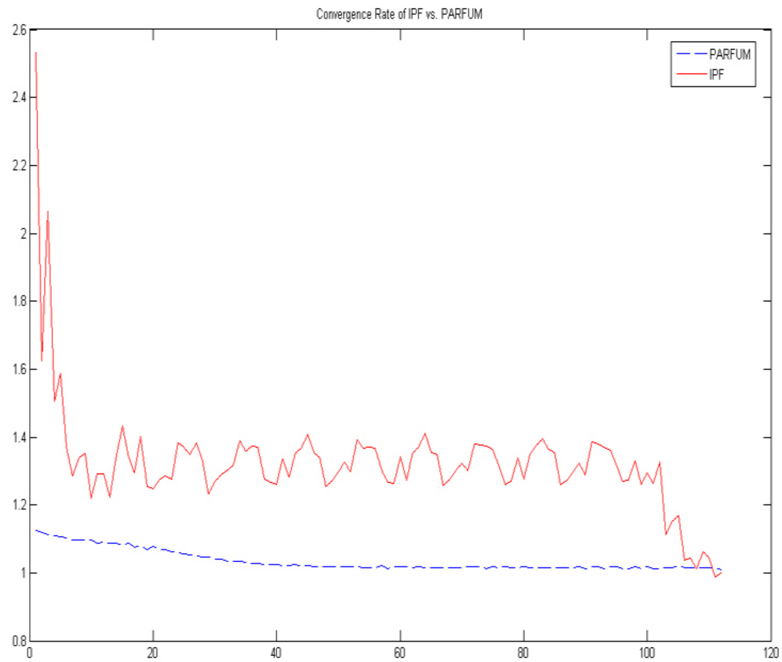

Fig. 6.2: Convergence rate of both iterative solvers

6.5 GCI vs. Scoring of Models

Fitting either one of the scores acquired from the models to the Global Competitiveness Index (GCI) would lead to a relative bad fit, because there is a clear distinction between the ranking from the GCI and the ranking of the scores. This can be emphasized by the following scatterplots and values of the R-squared.

Country	1	2	3	4	5	6	7	8	Score	Ranking
1	0	17	15	22	15	14	22	17	122	1
2	8	0	11	20	13	11	23	15	101	4
3	10	14	0	23	17	13	22	17	116	2
4	3	5	2	0	13	5	19	10	57	7
5	10	12	8	12	0	8	18	13	81	5
6	11	14	12	20	17	0	21	18	113	3
7	3	2	3	6	7	4	0	6	31	8
8	8	10	8	15	12	7	19	0	79	6

Tab. 6.13: Frequency matrix of 25 respondents

Country	ThurC	ThurB	B-TI	B-TII	IPF	PARFUM
1	0.6970	0.9715	0.2287	0.2337	0.6524	0.6542
3	0.6122	0.6573	0.1977	0.1854	0.6246	0.6253
6	0.5014	0.7840	0.1841	0.1867	0.6124	0.6134
2	0.3148	0.3083	0.1397	0.1325	0.5597	0.5599
5	-0.1160	-0.2068	0.0892	0.0997	0.4715	0.4699
8	-0.1524	-0.2363	0.0853	0.0895	0.4642	0.4631
4	-0.6647	-0.5412	0.0508	0.0400	0.3643	0.3643
7	-1.1923	-1.7364	0.0246	0.0268	0.2487	0.2490

Tab. 6.14: Scores of models with 25 respondents

The next thing that can be done is to look if the scores can be explained by a subset of the nine pillars of the GCI. The whole set of pillars can not be used in the regression exercise, because the dependent vector (scores) is of size eight. The subset should contain at most eight pillars (without) intercept and seven with. Before doing the regression the values of the pillars or standardized, because they have different units of measurement. The pillars are standardized by computing the Z-scores $\frac{x_i-\bar{x}}{\sqrt{Var(x)}}$, where $\bar{x}$ is the mean of the values. After standardizing the pillars the correlation matrix is computed to determine the dependency of the pillars. The correlation matrix of the pillars is given as follows.

From the product moment correlation matrix it may be concluded that there some of the pillars are correlated with each other. It may be noted that one can not draw conclusion from such a small dataset for each pillar, but the correlation coefficient seem to reflect the reality. For example better **Higher Education** is defined as the quality and quantity of higher education provided within an economy are critical for competitiveness, for preparing qualified staff for more complex roles in areas, such as production, marketing, management, and R&D. And **Innovation** is defined as a creation (a new device or process)

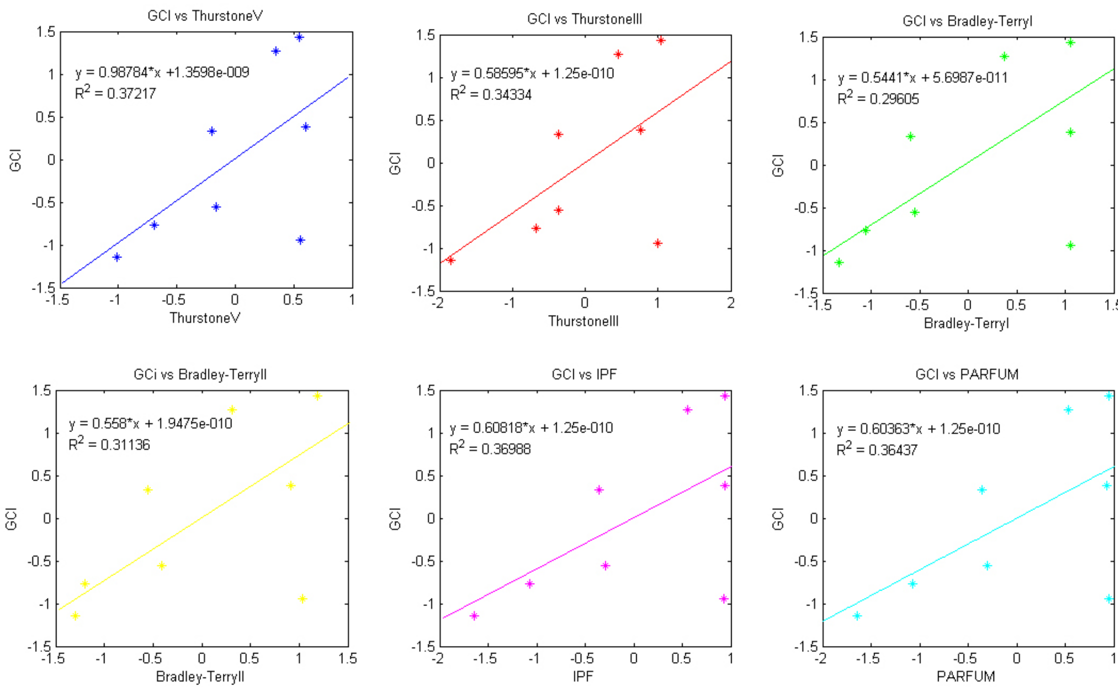

Fig. 6.3: GCI vs. Scores

resulting from study and experimentation. So a country with a high value of the pillar **Higher Education** will most likely have a high value of the pillar **Innovation**. The only correlation coefficient that seems odd is the correlation coefficient of the pillar **Macroeconomy** and the pillar **Health and Primary Education**. The correlation coefficient indicates that a if a country becomes macroeconomic more stable the health and primary education of that country will become degrade and this of course makes no sense.

Stepwise regression is used to find a subset of pillars that can best describe one of the scores obtained from one of the models. Stepwise regression can either start with a empty set and add a pillar to the set as long as the probability of seeing a *t-value*[2] from the regression coefficient from pillar is smaller than a given significance level α. Usually the value for α is equal to 0.05.

[2] The t statistic is the coefficient divided by its standard error. The standard error is an estimate of the standard deviation of the coefficient, the amount it varies across cases. It can be thought of as a measure of the precision with which the regression coefficient is measured. If a coefficient is large compared to its standard error, then it is probably different from 0.

ID	1	2	3	4	5	6	7	8	9
1	1.000	-0.167	0.069	-0.121	-0.744	-0.123	-0.016	-0.076	-0.119
2	-0.167	1.000	0.906	0.862	0.260	0.912	0.876	0.916	0.814
3	0.069	0.906	1.000	0.906	0.149	0.935	0.963	0.905	0.888
4	-0.121	0.862	0.906	1.000	0.358	0.963	0.972	0.869	0.922
5	-0.744	0.260	0.149	0.358	1.000	0.422	0.323	0.39	0.390
6	-0.123	0.912	0.935	0.963	0.422	1.000	0.980	0.968	0.912
7	-0.016	0.876	0.963	0.972	0.323	0.980	1.000	0.919	0.950
8	-0.076	0.918	0.905	0.869	0.398	0.968	0.919	1.000	0.820
9	-0.119	0.814	0.888	0.922	0.390	0.912	0.950	0.820	1.000

Tab. 6.15: Product moment correlation matrix from the pillars of WEF

6.5.1 Thurstone case C and The Pillars

The scores of the Thurstone case C can be described by at most three pillars and the three pillars that give the highest $multipleR-squared$ are given by the pillars one (**Macroeconomy**), six (**Higher Education and Training**) and seven (**Infrastructure**). The results are given as follows.

Coefficient	Value	Std. Error	t-value	$P(>\|t\|)$
β_0	0.0000	0.0866	0.0000	1.0000
β_1	0.6881	0.1099	6.2627	0.0033
β_6	2.2239	0.5509	4.0366	0.0156
β_7	-2.0907	0.5468	-3.8234	0.0187

Tab. 6.16: Regression results with Thurstone case C as dependent variable

where β_0 is the intercept.

$Residual\ standard\ error$: 0.2448 on 4 degrees of freedom
$Multiple\ R-Squared$: 0.9102
$F-statistic$: 13.51 on 3 and 4 degrees of freedom, the $p-value$ is 0.01466

6.5.2 Thurstone case B and The Pillars

The scores of the Thurstone case B can be described by at most six pillars. These six pillars are given by the pillars one (**Macroeconomy**), five (**Health and Primary Education**), six (**Higher Education and Training**), seven (**Infrastructure**), eight (**Institutions**) and nine (**Business Sophistication**). With seven pillars the multiple R-squared does not differ significantly from the multiple R-squared with six pillars and the statistical test can not be carried out. The $multipleR - squared$ has a value of 1 with seven pillars and 0.99999963513957679 with six pillars. Using the fact that the pillars are also highly correlated is also an argument to choose for six pillars instead of seven. The results of the regression are given as follows

Coefficient	Value	Std. Error	t-value	P(> $\|t\|$)
β_0	0.0000	0.0006	0.0000	1.0000
β_1	0.9932	0.0013	744.7285	0.0009
β_5	-0.2629	0.0011	-241.0856	0.0026
β_6	5.3118	0.0107	494.4195	0.0013
β_7	-3.2484	0.0089	-366.5106	0.0017
β_8	-1.1572	0.0042	-274.9266	0.0023
β_9	-0.5300	0.0027	-195.6201	0.0033

Tab. 6.17: Regression results with Thurstone case B as dependent variable

$Residual\ standard\ error$: 0.001598 on 1 degrees of freedom
$Multiple\ R - Squared$: 1
$F - statistic$: 456800 on 6 and 1 degrees of freedom, the $p - value$ is 0.001133

6.5.3 Iterative Solution of TheBradley-Terry Model and The Pillars

The scores of the iterative solution of the Bradley-Terry model can be described with the same pillars as for the scores of the Thurstone case C. The results of the regression are given as follows

Coefficient	Value	Std. Error	t-value	P(> $\|t\|$)
β_0	0.0000	0.1546	0.0000	1.0000
β_1	1.0964	0.1963	5.5856	0.0050
β_6	2.9999	0.9842	3.0479	0.0381
β_7	-2.7275	0.9769	-2.7919	0.0492

Tab. 6.18: Regression results with the iterative solution of the Bradley-Terry as dependent variable

$Residual\ standard\ error$: 0.4374 on 4 degrees of freedom
$Multiple\ R - Squared$: 0.8907
$F - statistic$: 10.86 on 3 and 4 degrees of freedom, the $p - value$ is 0.02157

6.5.4 The Least Square Solution of The Bradley-Terry Model and The Pillars

The scores of the least square solution of the Bradley-Terry model can be described with the same pillars as for the scores of the Thurstone case C. The results of the regression are given as follows

Coefficient	Value	Std. Error	t-value	P(> $\|t\|$)
β_0	0.0000	0.1601	0.0000	1.0000
β_1	1.0903	0.2032	5.3655	0.0058
β_6	3.2846	1.0189	3.2237	0.0322
β_7	-3.0141	1.0113	-2.9803	0.0407

Tab. 6.19: Regression results with the least square solution of the Bradley-Terry model as dependent variable

$Residual\ standard\ error$: 0.4528 on 4 degrees of freedom
$Multiple\ R-Squared$: 0.8829
$F-statistic$: 10.05 on 3 and 4 degrees of freedom, the $p-value$ is 0.0247

6.5.5 IPF and The Pillars

The scores of the non-parametric model, which is solved by IPF can be described with the same pillars as for the scores of the Thurstone case C. The results of the regression are given as follows

Coefficient	Value	Std. Error	t-value	P(> $\|t\|$)
β_0	0.0000	0.1377	0.0000	1.0000
β_1	1.1115	0.1748	6.3602	0.0031
β_6	3.5423	0.8763	4.0424	0.0156
β_7	-3.2941	0.8698	-3.7872	0.0193

Tab. 6.20: Regression results with IPF

$Residual\ standard\ error$: 0.3894 on 4 degrees of freedom
$Multiple\ R-Squared$: 0.9133
$F-statistic$: 14.05 on 3 and 4 degrees of freedom, the $p-value$ is 0.01366

6.5.6 PARFUM and The Pillars

The scores of the non-parametric model, which is solved by PARFUM can be described with the same pillars as for the scores of the Thurstone case C. The results of the regression are given as follows

Coefficient	Value	Std. Error	t-value	P(> $\|t\|$)
β_0	0.0000	0.1383	0.0000	1.0000
β_1	1.1117	0.1755	6.3330	0.0032
β_6	3.5089	0.8802	3.9866	0.0163
β_7	-3.2600	0.8736	-3.7316	0.0203

Tab. 6.21: Regression results with PARFUM

Residual standard error: 0.3911 on 4 degrees of freedom
Multiple $R-Squared$: 0.9126
$F-statistic$: 13.92 on 3 and 4 degrees of freedom, the $p-value$ is 0.0139

6.5.7 Remarks Regression with the Pillars

It seems that all the scores can be described by the pillars **Macroeconomy**, **Higher Education and Training** and **Infrastructure** except for the scores obtained from the Thurstone case III model. The only thing not satisfactory is the fact that the regression coefficient for the pillar **Infrastructure** is negative. Good infrastructure is usually needed for the development of a country. A country with bad infrastructure can most like not transport and facilitate *goods* and *people*, which in return will most likely be less competitive than a country with better infrastructure.

It seems odd that the scores of the Thurstone III case model can be described by six pillars, but the results of the regression coefficient are not. There is high correlation between the pillars seven and eight, seven and nine, and pillar one is negatively correlated with pillar five. The regression coefficient of pillar seven was negative for all the other models so if more pillars are added that are positive highly correlated with pillar seven then they will most likely be also negative. Pillar five is highly negative correlated with pillar one so its regression will most likely also be negative if the regression coefficient of pillar one is positive.

Finally the pillars of the WEF are an aggregated indicator of soft and hard data so it seems best to search for single indicators that are drivers of the concept Competitiveness.

6.6 Looking for Drivers of Competitiveness

The last objective of the study was to search for indicators that describe the concept competitiveness. Finding data for this exercise turned out to be more difficult than expected. The idea was to search the databases from the independent organization that gather data from countries. Most of the independent organizations did not include data for the eight selected countries. Only one indicator was found from the factbook of the **OECD** (Organisation for Economic Co-operation and Development)[3]. This indicator is the indicator **Evolution of GDP** of **Macroeconomic trends**. There may be more indicators available, but they were not found within the time constraint of this study. The following table shows the Z-scores of the scores of all models and the indicator **Evolution of GDP**.

Country	ThurV	ThurIII	B-TI	B-TII	IPF	PARFUM	GDP
1	0.5519	1.0363	1.0532	1.1898	0.9406	0.9462	-0.3438
3	0.6027	0.7677	1.0532	0.9100	0.9393	0.9330	-0.5375
6	0.5600	1.0032	1.0532	1.0371	0.9311	0.9481	2.1680
2	0.3488	0.4542	0.3685	0.3119	0.5500	0.3488	-0.0824
5	-0.1607	-0.3703	-0.5460	-0.4055	-0.2897	-0.2968	0.0956
8	-0.1985	-0.3708	-0.5960	-0.5478	-0.3565	-0.3526	0.0932
4	-0.6956	-0.6796	-1.0588	-1.1955	-1.0679	-1.0680	-1.3666
7	-1.0087	-1.8406	-1.3274	-1.3000	-1.6469	-1.6427	-0.0264

Tab. 6.22: Scores of models versus Evolution of GDP

Regression is not done with this single indicator only, because the ranking of this indicator differs from the ranking of all the scores. Ans also the goal of this study was to demonstrate the method.

6.7 Remarks study

First of all it must be stated that the respondents used in this study are not experts in the fields of assessing country's performance. The scores acquired by either one of the models are thereby not so creditable. This can be explained by the following example. Suppose a person is asked to choose between two unknown items, then this person may evaluate his or hers choice as follows. Choose the item randomly if absolute nothing is known about either of the item's attributes or if some attributes are know about both items choose the item for which its attributes give the highest satisfaction. There is nothing wrong with the last option, but when one searches for an aggregated index it is important that a set of persons is selected who are suppose to know a lot about the items

[3] The OECD groups 30 member countries sharing a commitment to democratic government and the market economy. With active relationships with some 70 other countries, NGOs and civil society, it has a global reach. Best known for its publications and its statistics, its work covers economic and social issues from macroeconomics, to trade, education, development and science and innovation.

selected for the study. Also must be kept in mind that the index obtained is subjective and it can only be used to gather people beliefs.

Second, because of the fact that the respondents are not real experts the indicators that are found via regression may or may not be the drivers of global competitiveness. Further more countries should be used or perhaps a set of union regions, e.g. the European Union, the Caricom, US, Arabic Countries (with a pact), and Asian countries (with a pact) to get a better view of the concept global competitiveness. How many more countries? Preferably all of the countries of the world, but then the number of paired will be too large. There are approximately 190 countries (depending on the sources used) so 17955 paired comparisons need to be carried out. And suppose by any chance these paired comparisons have been carried out the solvers of the non-parametric models will have a hard time of finding a solution.

Finally, this study did not deal with incomplete data or with data that could cause problems. All the proportions of (i, j) were between zero and one. The Thurstone models and the non-parametric model that uses IPF would not be able to proportions, if one of these proportions is equal zero or one. IPF would not converge, because the cell accompanied by the proportion one would be empty.

7. CONCLUSIONS AND FUTURE WORK

The study has investigated an alternative method to elicit a global competitiveness indicator over a set of countries, and to identify a set of indicators that are drivers of the concept global competitiveness.

7.1 The Alternative Method

The alternative method used is not new, but rather a new application of the alternative method. However two new alternative models were formulated and successfully used. The first model is the modified Bradley-Terry model, which uses the least square method instead of a iterative solver to solve the modeling assumption of Bradley-Terry. The second model is a non-parametric model, which either can be solved using constraint optimization or using iterative sample re-weighting. The solver that uses constraint optimization was not applicable for the research, because it could only handle a set with less than eight countries.

The non-parametric models are capable to determine dependency between alternatives. The original Thurstone model is intended to discover the dependency, but can not be solved using previously know methods. The non-parametric model abandoned the notion that the scores of the alternative follow a specific distribution.

7.2 The drivers

Regression can be a good tool to identify drivers of a given concept. In case of the study the concept was global competitiveness. The results for the regression were not so satisfactory using the nine pillars of the World Economic Forum **WEF**, but if one look at the pillars obtained as a driver closely one might conclude that macroeconomic stability, and higher education and training are an acceptable driver of global competitiveness. A country needs to have enough resources and sufficient knowledge to compete, e.g. a battle can not be won if there is not enough resources and know-how.

7.3 Future Work

The non-parametric models used are new and there will be still new things to discover about them. For example what are good convergence criteria's, what can be said of the following sum $\sum_{i=1}^{n} \sqrt{Var\left(u_i\right)}$, and could a different approach solve the non-parametric model that uses constraint optimization?

The research also started with creating a framework of eliciting stakeholders' preferences. Via the website people are able to input their preferences and these preferences are then saved into a centralized database. Researchers are then able to access this centralized database and use this database for their study.

In the future all the transactions that are made within the framework should be authenticated. Simple authentication is now used to store the results obtained by the respondents, but when researchers use the data an authentication and or validation process should be used. That is why in the program created researchers are not able to "create" respondents for a study (respondents most likely have a set of comparisons carried out).

APPENDIX

A. SOLVER CLASS USED IN STUDY

For each of the models to determine a scoring out of paired comparisons data there is written a solver class. However the solver classes all inherit from a base class called "Judgment". This class has the implementation of all standard functions like for example computing the coefficient of concordance. The C-Sharp for this code is is given as follows.

```
using System;
using System.Collections;
using System.IO;
using System.Text;
using System.Xml;
using System.Xml.Serialization;
using ThesisII.Core;
using ThesisII.Solvers.Functions;

namespace ThesisII.Solvers
{
    public abstract class Judgment
    {
        #region Properties
        private Matrix _F,_P,_X;
        private Stimuli _stimuli;
        private Experts _experts;
        private Vector _nk;
        private Vector _fk;
        private double _CoefficientOfAgreement;
        private double _CoefficientOfConcordance;
        private double _ProbConcordance;
        private double _ProbAgreement;
        #endregion

        #region Constructors
        public Judgment()
        {
        }

        public Judgment(Stimuli stimuli, Experts experts)
        {
            this.stimuli = stimuli;
            this.experts = experts;
            this.init(true);
        }
        #endregion

        #region Properties
        public Matrix F
        {
            get
            {
                return _F;
            }
            set
            {
                _F = value;
            }
        }
        [XmlIgnore]
        public Matrix P
        {
            get
            {
                return _P;
            }
            set
            {
                _P = value;
            }
        }
        [XmlIgnore]
```

```
public Matrix X
{
    get
    {
        return _X;
    }
    set
    {
        _X = value;
    }
}

public Stimuli stimuli
{
    get
    {
        return _stimuli;
    }
    set
    {
        _stimuli = value;
    }
}
[XmlIgnore]
public Experts experts
{
    get
    {
        return _experts;
    }
    set
    {
        _experts = value;
    }
}

public Vector Nk
{
    get
    {
        return _nk;
    }
    set
    {
        _nk = value;
    }
}

public Vector Fk
{
    get
    {
        return _fk;
    }
    set
    {
        _fk = value;
    }
}

public double CoefficientOfAgreement
{
    get
    {
        return getCoefficientOfAgreement();
    }
    set
    {
        _CoefficientOfAgreement = value;
    }
}

public double CoeffiecientOfConcordance
{
    get
    {
        return getCoeffientOfConcordance();
    }
    set
    {
        _CoefficientOfConcordance = value;
    }
}

public double PConcordance
{
    get
    {
        return probConcordance();
    }
    set
```

```
        {
            _ProbConcordance = value;
        }
    }

    public double PAgreement
    {
        get
        {
            return probAgreement();
        }
        set
        {
            _ProbAgreement = value;
        }
    }
    #endregion

    #region Methods
    public void init(bool firt_time)
    {
                    int size =this.stimuli.stimuli.Count;
        if (firt_time)
        {
            this.F = new Matrix(new double[size, size]);
            this.Fk = new Vector(new double[size]);

            foreach (Expert expert in this.experts.experts)
            {
                if (this.experts.alpha == 0D)
                {
                    if (expert.Selected)
                    {
                        this.F = this.F + expert.Matrix;
                        this.Fk = this.Fk + expert.fk;
                    }
                }
                else
                {
                    if (expert.probTriad() > (1 - this.experts.alpha)
                   && expert.Triads > -1)
                    {
                        this.F += expert.Matrix;
                        this.Fk += expert.fk;
                        expert.Selected = true;
                    }
                }
            }

            this.Nk = new Vector(new double[size]);
            this.P = new Matrix(new double[this.F.rows, this.F.cols]);

            for (int i = 0; i < this.P.rows; i++)
            {
                for (int j = i + 1; j < this.P.cols; j++)
                {
                    this.P[i, j] = this.F[i, j] / this.Fk[i];
                    this.P[j, i] = 1 - P[i, j];
                }
            }

            this.X = new Matrix(new double[this.F.rows, this.F.cols]);

            for (int i = 0; i < this.X.rows; i++)
            {
                for (int j = 0; j < this.X.cols; j++)
                {
                    if (i != j)
                    {
                        this.X[i, j] =
                        SpecialFunction.getInvCDF(this.P[i, j], true);
                    }
                    else
                    {
                        this.X[i, j] = SpecialFunction.getInvCDF(-1, true);
                    }
                }
            }

            for (int i = 0; i < this.X.rows; i++)
            {
                for (int j = 0; j < this.X.cols; j++)
                {
                    if (i != j)
                    {
                        if (!double.IsInfinity(this.X[i, j]))
                        {
                            this.Nk[j]++;
                        }
                    }
                    else
```

```
                    {
                        this.Nk[j]++;
                        this.X[i, j] = 0D;
                    }
                }
            }
        }
        else
        {
            this.Nk = new Vector(new double[size]);
            this.P = new Matrix(new double[this.F.rows, this.F.cols]);

            for (int i = 0; i < this.P.rows; i++)
            {
                for (int j = i + 1; j < this.P.cols; j++)
                {
                    this.P[i, j] = this.F[i, j] / this.Fk[i];
                    this.P[j, i] = 1 - P[i, j];
                }
            }

            this.X = new Matrix(new double[this.F.rows, this.F.cols]);

            for (int i = 0; i < this.X.rows; i++)
            {
                for (int j = 0; j < this.X.cols; j++)
                {
                    if (i != j)
                    {
                        this.X[i, j] =
                        SpecialFunction.getInvCDF(this.P[i, j], true);
                    }
                    else
                    {
                        this.X[i, j] = SpecialFunction.getInvCDF(-1, true);
                    }
                }
            }

            for (int i = 0; i < this.X.rows; i++)
            {
                for (int j = 0; j < this.X.cols; j++)
                {
                    if (i != j)
                    {
                        if (!double.IsInfinity(this.X[i, j]))
                        {
                            this.Nk[j]++;
                        }
                    }
                    else
                    {
                        this.Nk[j]++;
                        this.X[i, j] = 0D;
                    }
                }
            }
        }
    }

    public double getCoeffientOfConcordance()
    {
                    int size = this.stimuli.stimuli.Count;
        double sum = 0D;
        double avg = 0D;
        double deviation = 0D;
        double denum = (1/12.0)*
        Math.Pow(this.Fk[0],2)*((Math.Pow(size,3) - size));
        Vector totalPref = new Vector(size);
        foreach (Expert expert in this.experts.experts)
        {
            totalPref = totalPref + expert.A;
        }
        for (int i = 0; i < totalPref.Length; i++)
        {
            sum = sum + totalPref[i];
        }
        avg = sum / totalPref.Length;
        for (int i = 0; i < totalPref.Length; i++)
        {
            deviation = deviation + Math.Pow(totalPref[i] - avg, 2);
        }

        return deviation / denum;
    }

    public double probConcordance()
    {
        double sum = 0D;
        double avg = 0D;
        double deviation = 0D;
```

```
            int size = this.stimuli.stimuli.Count;
            double denum = (1 / 12.0) * (this.Fk[0]) * (size)*(size + 1);
            Vector totalPref = new Vector(size);
            foreach (Expert expert in this.experts.experts)
            {
                totalPref = totalPref + expert.A;
            }
            for (int i = 0; i < totalPref.Length; i++)
            {
                sum = sum + totalPref[i];
            }
            avg = sum / totalPref.Length;
            for (int i = 0; i < totalPref.Length; i++)
            {
                deviation = deviation + Math.Pow(totalPref[i] - avg, 2);
            }

            return SpecialFunction.chisq(size -1,deviation / denum);
        }

        public double getCoefficientOfAgreement()
        {
            double sum = 0;
            int size = this.stimuli.stimuli.Count;
            for( int i = 0; i < size;i++)
            {
                for (int j = 0; j < size;j++ )
                {
                    if (i != j)
                    {
                        sum = sum + Binomial((int)this.F[i, j], 2);
                    }
                }
            }

            return ((2 * sum) / (Binomial((int)Fk[0], 2) * Binomial(size, 2))) - 1 ;
        }

        public double probAgreement()
        {
            double sum = 0D;
            int size = this.stimuli.stimuli.Count;
            for (int i = 0; i < size; i++)
            {
                for (int j = 0; j < size; j++)
                {
                    if (i != j)
                    {
                        sum = sum + Binomial((int)this.F[i, j], 2);
                    }
                }
            }
            double uprime = 4 * (sum - ((Binomial((int)Fk[0], 2)
            * Binomial(size, 2)))
            * (this.Fk[0] - 3) / (2 * this.Fk[0] - 4)) / (this.Fk[0] - 2);
            double v = (Binomial(size, 2) * this.Fk[0] * (this.Fk[0] - 1)) /
            ((this.Fk[0] - 2) * (this.Fk[0] - 2));
            return SpecialFunction.chisq(v, uprime);
        }

        public int Binomial(int n, int k)
        {
            int result = 1;
            for (int i = 1; i < k + 1; i++)
            {
                result = result * (n - i + 1) / i;
            }
            return result;
        }

        #endregion
    }
}
```

BIBLIOGRAPHY

[1] Kenneth J. Arrow. *Social Choice and Individual Values.* John Wiley & Sons, Inc., 1951.

[2] Kenneth J. Arrow. *Essays in the Theory of Risk-Bearing.* Markham Publishing Company, Chicago, 1971.

[3] R. Bradley. *Some Statistical Methods in Taste Testing and Quality Evaluation.* Biometrica, 1953.

[4] R. Cooke. Discrete choice with probabilistic inversion, 2006.

[5] Roger M. Cooke. *Experts in Uncertainty Opinion and Subjective Probability in Science.* Oxford University Press, Inc., New York, 1991.

[6] H. A. David. *The Method of Paired Comparisons.* Charles Griffin, 1963.

[7] L. Ford. *Solution of a Ranking Problem from Binary Comparisons.* Amer. Math. Monthly, 1957.

[8] World Economic Forum. World economic forum - our orginization, 2003. [Online; last updated 18-July-2006].

[9] M. Kendell. *Rank Correlation Methods.* Charles Griffin & Co. Limited, London, 1962.

[10] Xavier Sala-i-Marti Michael E. Porter, Klaus Schwab and Augosto Lopes-Carlos. *The Global Competitiveness Report 2004-2005.* Palgrave Macmillan, New York, 2004.

[11] F. Mosteller. Remarks on the method of paired comparisons: Iii a test of significance for paired comparisons when equal standard deviations and equal correlations are assumed, 1951.

[12] Arie H. Havelaar-Ine van der Fels R. Cooke, Maarten Nauta. Probabilistic inversion for chicken processing lines, 2005.

[13] L. Savage. *The Foundations of Statistics.* John Wiley & Sons, 1954.

[14] S. Siegel and N.J. Castellan Jr. *Nonparametric Statistics for The Behavioral Sciences.* McGraw-Hill, New York, 1988.

[15] Alfredo Di Tillio. Subjective expected utility in games, 2004.

[16] Warren S. Torgerson. *Theory and Methods of Scaling.* John Wiley & Sons, Inc., New York, 1963.

[17] Wikipedia. World economic forum — wikipedia, the free encyclopedia, 2006. [Online; accessed 22-July-2006].

[18] J. M. Wooldridge. *Introductory Econometrics.* Thomson Learning, Ohio, 2002.

European Commission
Directorate-General Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Address: Via E. Fermi, 1 – 21020 Ispra (VA) Italy
E-Mail: Stefano.tarantola.jrc.it
Tel.: +39 0332 789928
Fax: +39 0332 785733

http://stat-ind.cec.eu.int
http://www.jrc.cec.eu.int

EUR XXXXX LL
Luxembourg: Office for Official Publications of the European Communities

Printed in the Netherlands

European Commission

EUR XXXXX LL – DG Joint Research Centre, Institute for the Protection and Security of the Citizen

Luxembourg: Office for Official Publications of the European Communities
20006 – 75 pp. 21.59 x 27.94 cm
Scientific and Technical Research series

Abstract

Statisticians often argue that all the work done during the composition of a composite indicator, to collect data and to edit this data, seems to be ”hidden” or ”wasted”. On the other hand stakeholders like the idea that a composite indicator is capable of summarizing complex and or difficult concepts into one single number. The research will start by changing the process of obtaining a composite indicator. At first a composite indicator will be obtained by eliciting stakeholders' preferences. After obtaining the composite indicator the work that statisticians put into collecting and editing data, is used to identify the drivers behind the concept of the composite indicators, using regression. The methods to elicit stakeholders' preferences are not new and are known as paired comparison analysis. However new models are used to obtain an index out of the paired comparison data. These new models make use of Probabilistic Inversion and use IPF or PARFUM to solve them. Also these new models make no assumptions about the distribution of the scores of the alternatives as the traditional methods do, but obtain better variances estimated and covariance.

Keywords: Composite indicators, stakeholders' preferences, regression, paired comparison analysis, Probabilistic Inversion, IPF, and PARFUM.

The mission of the Joint Research Centre is to provide customer-driven scientific and technical support for the conception, development, implementation and monitoring of European Union policies. As a service of the European Commission, the JRC functions as a reference centre of science and technology for the Union. Close to the policy-making process, it serves the common interest of the Member States, while being independent of special interests, whether private or national.