Computational Intelligence for Measuring Macro-Knowledge Competitiveness

AHMAD Q. AL-SHAMI

A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of

Doctor of Philosophy

August 2013

To my beloved mother with great gratitude for her lifelong sacrifice for the whole family. Undoubtedly, without her prayers, endless love and encouragements this thesis would have been impossible. Thank you mom for every beautiful thing you did to make our life beautiful.

Acknowledgements

First and foremost, all praise and thanks are due to the almighty Allah, for his blessings, help and mercy along this journey and I render this and future work for His sake.

I owe my sincere appreciations and gratitude to my respectable director of studies, Dr. Ahmad Lotfi, for his guidance over the last four years. His office was always open for me for any inquiry despite his many responsibilities. Also, his patience, relentless effort and extensive suggestions around my work were indispensable to the accomplishment of this research. I would like to extend a huge thank to Dr. Simeon Coleman for his advice, support, feedback and knowledge throughout the development of my thesis. I also thank Dr. Eugene Lai, for his support in the early stage of this project. Acknowledgement is also due to Professor Petr Dostál, Brno University of Technology, Czech Republic, who gave me valuable suggestions in our discussion during his visit to Nottingham.

Moreover, I would like to show appreciation to the financial support of the School of Science and Technology at Nottingham Trent University to participate in various international conferences. The data sets provided by the IMD Business School and the International Telecommunication Union which partially utilised in this research is also gratefully acknowledged.

Thanks are extended to my extended family: mother, mother and father in law, brothers and brothers in law. Your prayers and faith have inspired and encouraged me to continue with this study. In particular, I owe a huge debt of gratitude to my dear brothers Waleed and Abdullah for huge support. Without your precious guidance, insights, and generosity, the graduate study would have been very difficult. Thanks also should gone to friends and colleagues in the School of Science and Technology / The Computing and Informatics Department in Nottingham Trent University. In addition, I would like to pay tribute to my family for their dedications, support, sympathy, and patience throughout the journey of this study. In particular, I would like to say big thanks to my wife Fatemah Alzahraa' and our courageous son Humam and lovely daughters, Zaynab and Aya; your wonder and love have always inspired me during the happiness and sorrow of the last four years.

Ahmad Q. Al Shami
August 2013

Abstract

The aim of this research is to investigate the utilisation of Computational Intelligence methods for constructing Synthetic Composite Indicators (SCI). In particular for delivering a Unified Macro-Knowledge Competitiveness Indicator (*UKCI*) to enable consistent and transparent assessments and forecasting of the progress and competitiveness of Knowledge Based Economy (KBE). SCI are assessment tools usually constructed to evaluate and contrast entities performance by aggregating intangible measures in many areas such as economy, education, technology and innovation. SCI key value is inhibited in its capacity to aggregate complex and multi-dimensional variables into a single meaningful value. As a result, SCIs have been considered as one of the most important tools for macro-level and strategic decision making. Considering the shortcomings of the existing SCI, this study is proposing an alternative approach to develop Intelligent Synthetic Composite Indicators *(iSCI)*. The suggested approach utilizes Fuzzy Proximity Knowledge Mining technique to build the qualitative taxonomy initially, and Fuzzy c-mean is employed to form the new composite indicators.

To illustrate the method of construction for the proposed *iSCI*, a fully worked application is presented. The presented application employs Information and Communication Technology (ICT) real variables to form a new unified ICT index. The weighting and aggregation results obtained were compared against classical approaches namely Vector Quantisation and Principal Component Analysis, Factor Analysis and the Geometric mean to weight and aggregate synthetic composite indicators. This study also compares and contrasts Optimal Completion

Strategy and the Nearest Prototype Strategy to substitute missing values. The validity and robustness of the techniques are evaluated using Monte Carlo simulation.

The developed *iSCI* concept is generalised to build the suggested *UKCI* which ultimately is equipped with short-term forecasting capabilities. This achieved by a hybridised model consisting of Artificial Neural Networks and Panel Data: Time Series Cross Sectional to predict and forecast the competitiveness of KBE. The proposed model has the capability of forecasting and aggregating seven major KBE indicators into a unified meaningful map that places any KBE in its league even with limited data points. The Unified Knowledge Economy Forecast Map reflects the overall position of homogeneous knowledge economies, and it can be used to visualise, identify or evaluate stable, progressing or accelerating KBEs. In order to show the value added by the new development techniques, the *UKCI* is applied to fifty-seven countries initially, then expanded to include the Middle East and North Africa (MENA) region as a special case study. In total seventy-three countries were included, that are representative of developed, developing and underdeveloped economies. The final and overall results obtained, suggest novel, intelligent and unbiased results compared to traditional or statistical methods when building, not only the *UKCI*, but for any future composite indicator for many other fields.

Publications

The following publications have been published as a direct result of this thesis:

Refereed Journal Papers

Ahmad Al Shami, Ahmad Lotfi and Simeon Coleman "Intelligent Synthetic Composite Indicators with Application," **Soft Computing:** Volume 17, Issue 12 (2013), Page 2349-2364, Springer Berlin Heidelberg, DOI: 10.1007/s00500-013-1098-3, ISSN: 1432-7643.

Ahmad Al Shami, Ahmad Lotfi, Simeon Coleman and Petr Dostál, "Unified Knowledge Based Economy Forecasting," **Journal of Technological Forecasting & Social Change**, Elsevier (Accepted subject to revision).

Refereed Conference Papers

Ahmad Al Shami, Ahmad Lotfi and Simeon Coleman, "Towards Developing 3rd Generation Intelligent Synthetic Composite Indicators," UKCI 2012 IEEE 12th Annual Workshop on Computational Intelligence, September, 5-7, 2012 - Edinburgh, United Kingdom.

Ahmad Al Shami, Ahmad Lotfi and Simeon Coleman, "Unified Knowledge Based Economy Neural Forecasting Map," WCCI 2012 IEEE World Congress on Computational Intelligence (IJCNN), June, 10-15, 2012 - Brisbane, Australia, pp. 2008-2015. DOI: 10.1109/IJCNN.2012.6252645, E-ISBN : 978-1-4673-1489-3, Print ISBN: 978-1-4673-1488-6.

Ahmad Al Shami, Ahmad Lotfi, Eugene Lai, & Simeon Coleman, S. 2011, "Unified knowledge economy competitiveness index using fuzzy clustering model," IEEE Symposium Series on Computational Intelligence (SSCI 2011) / (CIFEr), Paris, France, April 11-15, 2011. DOI: 10.1109/CIFER.2011.5953563, Print ISBN: 978-1-4244-9933-5.

Ahmad Al Shami, Ahmad Lotfi, Eugene Lai, & Simeon Coleman, "Forecasting Macro-Knowledge Competitiveness; Integrating Panel Data Analysis and Computational Intelligence," In Proceedings of the 31st International Symposium on Forecasting (ISF 2011), Prague, Czech Republic, June 26-29, 2011.

Contents

List of Figures

List of Tables

Nomenclature

Roman Symbols

AHP Analytic Hierarchy Processes

AM Arithmetic Mean

ANFIS Adaptive Neuro-Fuzzy Inference System

ANN Artificial Neural Networks

AP Accelerating Progress

AR Additive Rules

ARIMA AutoRegressive Integrated Moving Average

ARMA AutoRegressive Moving Average

ARMAX AutoRegressive Moving Average with exogenous inputs

BAP Budget Allocation Processes

BOD Benefits of the Doubt

CA Cluster Analysis

CI Computational Intelligence

COA Conjoint Analysis

DEA Data Envelopment Analysis

ED Euclidean Distance

EM Expectation Maximisation

EW Equal Weighting

FA Factorial Analysis

FCM Fuzzy c-Means

FEM Fixed Effect Model

FIS Fuzzy Inference Systems

FPKM Fuzzy Proximity Knowledge Mining

GCI Global Competitiveness Index

GII Global Innovation Index

GM Geometric Mean

GMA Geometric Mean Aggregation

HDI Human Development Index

ICT Information Communication and Technology

IDI ICT Development Index

IEWB Index of Economic Well-Being

IMD International Institute for Management Development

INS INSEAD Business School

iSCI Intelligent Synthetic Composite Indicators

ITU International Telecommunication Union

KBE Knowledge-Based Economy

KE Knowledge Economy

KEI Knowledge Economy Index

KM Knowledge Management

KMS Knowledge Management System

KSOC Knowledge Societies

LA Linear Aggregation

LBG VQ Linde Buzo and Gray Vector Quantization

LMREG Linear Multiple Regression

LSDV Least Square with Dummy Variables

MCMC Markov Chain Monte Carlo

MENA Middle East and North Africa

MI Multiple Imputation

MP Moving Progress

MS Mean Substitution

NF1 Naïve Forecast 1

NN Nearest Neighbour

NP Neutral Progress

NPS Nearest prototype Strategy

NRI Networked Readiness Index

OCS Optimal Completion Strategy

OECD Economic Co-operation and Development

PA Public Administration

PCA Principal Components Analysis

PDA Panel Data Analysis

PDS Partial Distance Strategy

REG Regression

REM Random Effect Model

SCI Synthetic Composite Indicators

SDI Sustainable Development Index

SFC Soft Focused Crawler

SVM Support Vector Machines

TSCS Time Series, Cross Sectional

UCM Unobserved Components Models

UKCI Unified Knowledge Competitiveness Indicator

UN United Nation

VAF Variance Accounted For

VCM Variance Components Model

VQ Vector Quantization

WB World Bank

WCY World Competitiveness Yearbook

WDS Whole Data Strategy

WEF World Economic Forum

Chapter 1

Introduction

Knowledge and information are nowadays regarded as two of the key drivers of a nations progress and status, sometimes more so than wealth, land, capital or labour. Raw facts or 'data' can be considered as the basic building blocks which lead to the formation of information and development of knowledge. Information is data-in-context and it refers to the collection and manipulation of raw facts/data which is then organised in such a way that they have meaningful value beyond the facts. The functional relationships between data, information and knowledge are often interpreted as a pyramid comprising of three levels: data at the bottom, knowledge at the top with information as an intermediate stage. Tools and techniques are available to facilitate the progression from data (level 1) to Information (level 2). The progression from information (level 2) to knowledge (level 3) involves complex and indirect processes. For example, reading an article enables the reader to gather information about a chosen topic of interest, but developing knowledge from the gathered information is often a lengthy process requiring relevant experience, critical evaluation/reflective analysis and some prior knowledge. In some simple cases, it may be possible to develop knowledge directly from data without following the hierarchy.

In general, the main ingredients for creating new knowledge are data, information and time. It is equally important to appreciate that knowledge is required to make sense of data and information (i.e. to make them interpretable). A systematic procedure for distinguishing, gathering, saving, and converting information into knowledge is known as a Knowledge Management System (KMS) .

There are mixed definitions for knowledge, as it is viewed as an abstract concept and its potential can only be realised when it is used to some end. According to Stair and Reynolds (2007) knowledge is a blend of experience, values, information-in-context, and insight, while Knowledge Management (KM) refers to the process of comprehending, comparing, judging, remembering, and reasoning.

The application of the KM concept by nations has led to the development of Knowledge-Based Economies (KBEs). A KBE has two key strategies: first, it focuses on developing the means (hardware and software) to enable low-cost, high-speed electronic connectivity for data transfer; second, it aspires to produce a greater stock of highly skilled human capital to support and expand the prosperity of a nation. Nations that claimed to have a KBE see themselves as learning societies pursuing a policy of continuous improvement in their knowledge assets (Bontis et al., 1999). Furthermore, utilising knowledge as the key driver for prosperity and growth on a micro level leads to the creation of Knowledge Societies (KSOC), which relies on the production, distribution, and use of knowledge as the main driver of growth, wealth creation, and employment across all sectors (Quantumiii, 2011). This means that a nation must be in a constant process of change and adaptation to the new economic realities. In this new economy, it is evident that the key to success is 'knowledge'. For governments they have to invest in, and develop knowledge workers, which also means that government budgets should assign capital for education, skills and expertise development. (Riley, 2003). More recently, the term Knowledge Economy (KE) has been used to indicate whether a nation has an ability to create and export "expert" knowledge, thus enabling it to be seen as economically innovative (Michael, 2010).

Knowledge-based economies use a number of measures to inform themselves of their achievements. To this end, a range of indicators have been developed by highly regarded and reputable organisations including the United Nation (UN), the World Bank (WB), the World Economic Forum (WEF), the International Institute for Management Development (IMD), INSEAD business school (INS), the International Telecommunication Union (ITU) and many more. These indicators have been used by organisations including government agencies, aid agencies and research institutions to assess the progress of a nation or nations in the context of KBE or KE. However, these indicators are non-uniform, subjective and yield different scores and ranking depending on the nature and type of assessments (Archibugi et al., 2009). Moreover, most of these indicators report past performance, and it does not predict where a certain KBE is heading giving all known elements. To minimise any potential confusion to decision makers, a new and intelligent methods for developing progress or sustainability indicators in demand. Such methods should have the capacity to aggregate and forecast multi-dimensional and non-linear vectors within a limited time frame. This will encourage a consistent application of judgement and evaluation of performance, when developing and assessing a nation's KBE competitiveness using the published indicators.

1.1 Defining Terminology

When it comes to measuring the KBE, different indicators use different terms such as: progress, innovation, potential, advancement and development of a KBE; in an attempt to show the accuracy and distinct measurement. However, these words are multi-dimensional in nature, and they could be understood differently by different individuals which makes forming and using a unified and comprehensive meaning difficult. Nevertheless, looking up the definitions to these words will reveal that these words form an ontological relationship to each other. For example consider the definitions for the above mentioned terms as defined in Oxford Dictionary.

- **advancement:** a development or improvement,
- **progress:** a development towards an improved or more advanced condition,
- **development:** a specified state of growth or advancement, also it is defined as a new and advanced product or idea,
- **innovation:** making changes by introducing new ideas, or products,
- **competitiveness:** having a strong desire to be more successful than others, and

- **potential:** having or showing the capacity to develop into something in the future, also abilities that may lead to future success.

By connecting the similar words, one can clearly find the mentioned ontological relationship which circles around the meaning of well-being, welfare, happiness and prosperity of a nation on a macro scale; and it is no surprise to find indicators named as well-being index such as the Index of Economic Well-Being (IEWB). (Osberg and Sharpe, 2002). Such indicators are designed to measure things like how the community is progressing on increasing the wealth of its citizens, cleaning the environment, fostering innovations, developing education, leisure and culture (Osberg and Sharpe, 2005). The United Kingdom has introduced a new well-being index, to admit that GDP was an "imperfect way" of evaluating the country's development and to devise ways of measuring well-being and innovations (Self et al., 2012). Hence, this research will use any of the above mentioned terms interchangeably to cover the more general concept of measuring the well-being and competitiveness of a KBE on a macro level.

1.2 Research Motivations

Lately, information and knowledge are substituting energy and funds as the main source for nations prosperity, exactly as the former two substituted labour and land about 220 years back. Furthermore, technological advancements in the 21st century have changed the rules of wealth creating labour from tangibly-centred to knowledge-centred, where information and knowledge are becoming the main ingredients of high production (Botha, 2007). Measuring the competitiveness of KBE has become an important issue in recent years both for the public and private sectors. This implies investigative and evaluation techniques into the existing indicators used to measure the competitiveness of KBEs, which allow close attention to the underlying methods that constitute the making and formulation of Synthetic Composite Indicators (SCI). SCI are assessment tools, usually constructed to evaluate and contrast country performance, by aggregating abstract issues in many areas such as the economy, education, technology and innovation.

Nowadays many international organisations are focussing their attention on how to recognise, measure and promote developments in nations to improve the quality of lives, individually and collectively (OECD, 2008b). In order to achieve this growing trend, many statistical based composite indicators are developed. These indicators are qualitative and quantitative measures derived from a series of observed facts that can reveal relative position of a nation in a given area of progress (Saltelli et al., 2012). However, these indicators suffer many shortcomings, as they generate different ranking and scores depending on the nature and type of assessments, even though most of these indicators use the same variables and statistical techniques. For example, the Information Communication and Technology (ICT) Development Index (IDI), developed by the ITU, would include three variables (adult literacy rate, secondary enrolment ratio and tertiary enrolment ratio) to measure the level of ICT skills, these same variables would be used differently and under completely different pillar (category or basket) title by another index.

Due to the lack of availability of some information from a specific nation or during a specific period of time, there will be missing information. Another disagreement in ranking nations happens because of the different methods used to substitute missing values. For example, some organisations use the nearest neighbour imputation method, while others use mean or zero substitutions, and some use multiple imputation or expectation maximisation. Furthermore, the major problem is the difference in the methods used to weight and aggregate an index variables. These disagreements usually stem from the subjectivity or the opinions of the consulted "experts" who usually devise the methods to be used for building the framework, impute missing values, weight and aggregate the SCI variables. This and other controversies are also extensively explained by Tarantola et al. (2006) and Trebilcock and Prado (2011).

The challenges for this research in accordance to the above environment is to: introduce and investigate the use of Computational Intelligence (CI) techniques to learn KBEs behaviour using limited data sets, devise suitable and non-biased weighting and aggregation methodology, investigate and choose proper method for missing data problem. All are potentially critical where statistical methods has failed to deliver, specially, for underdeveloped and developing economies.

Many if not all developers of composite indicators relax the fact that these indices should represent and measure realistic events, and a set of synthetic aggregated indicators is not the reality, but it is basically an informative model of it (Saisana and Munda, 2008), therefore it is crucial to use a discipline that can construct better model of reality! One of the major insights of CI methods such as fuzzy logic is that many concepts are better defined by words than by mathematics, and fuzzy logic and its graded membership provide a discipline that can construct better model of reality (Cherchye and Kuosmanen, 2004). Hence, one of the main questions to be addressed in this thesis; is it possible to employ computational intelligence techniques to create a brand new unified and intelligent knowledge competitiveness framework? It is envisaged that the framework thus developed will contribute to the development of a computationally intelligent macro-knowledge framework and intelligent SCIs suitable for countries with different cultural, socio-economic, and technical conditions. This is expected to assist such developing economies in establishing, measuring, monitoring and forecasting suitable KBE.

In particular, this research study attempts to address the following:

- Why there are many different and separate views on KBE and competitiveness?
- What are the implications of these differences?
- Can computational intelligence offer an alternative technique to the measurement and forecast of KBE progress and competitiveness?
- Is it possible to unify existing KBE measures in a universally acceptable manner?
- What are the implications of using such techniques?

To answer the above questions, the aims and objectives of this research need to be expanded.

1.3 Aims and Objectives

The main argument and aim of this study is that given the different views about the nature of KE and competitiveness, the abundance of indicators and the confusion they create to the decision makers, a more intelligent, flexible and universally acceptable measure of the constituent elements contributing to KBE competitiveness can be better selected, weighted, aggregated and forecasted through the adoption of computationally intelligent approaches. The present research therefore will employ and assess such approaches to discover the utilisation of Computational Intelligence (CI) methods for constructing Synthetic Composite Indicators (SCI). In particular for delivering an intelligent qualitative taxonomy as a theoretical framework for making a Unified Macro-Knowledge Competitiveness Indicator (*UKCI*) to enable consistent and transparent assessments and forecasting of the progress and competitiveness of KBE. In this thesis, the focus is on whether different CI methodologies to build SCIs, would lead to different results. The use of CI techniques to build the quantitative side of a new SCI includes the use of Fuzzy Proximity Knowledge Mining (FPKM) methodology for the purpose of devising a non-biased, novel and intelligent method to create a new SCI taxonomy. This research also aims to fill the gap where existing KE indicators have failed. A contemporary and unified macro-knowledge epistemology is proposed, where many new factors such as intellectual capital and competitiveness would constitute a major ingredient for a reliable KBE measurement. Such new view would give credit to the efforts made by emerging, competitive and vibrant nations, which existing KBE indicators discounts.

The proposed methods to construct the *UKCI* will be applied to fifty-seven countries initially, then expanded to include the MENA region countries as a special case study. In total seventy three countries will be included, that are representative of developed, developing and underdeveloped economies. The *UKCI* will be evaluated on two levels: from a quantitative point of view and from real case study application in order to show the value added by the new development techniques and measure. The validity and robustness of all techniques are evaluated using Monte Carlo simulation. Finally, the *UKCI* will be subjected to a number of uncertainty and sensitivity analyses. It is envisaged that the KMS thus developed is capable of evaluating, measuring, describing, forecasting and analysing the main issues that affect knowledge economies on a macro level.

To accomplish the aim of this research, the following objectives are identified:

- To develop understanding and to critically evaluate the current and existing position of the KBEs by studying the available measurements and tools of utilisation issued by the global indicators.

- To propose an alternative method to the measurement of KBE and competitiveness, to integrate the strengths and resolve the shortcomings of the assessed techniques.

- To coin a novel CI technique as a a non-biased way to create the qualitative taxonomy of future SCIs. Fuzzy Proximity Knowledge Mining (FPKM) technique is proposed for this purpose. The suggested FPKM consists of two major steps: Focused web mining using Soft Focused Crawler (SFC), and fuzzy text matching using Wagner-Fischer dynamic programming algorithm for computing the Levenshtein or 'edit distance'. The suggested taxonomy will serve as a non-biased, novel and intelligent method for inclusion/exclusion and unifications of empirical variables to establish significant, consistent and sound SCI theoretical framework.

- To establish an intelligent and universally acceptable KBE measurement indicator; a number of analysis methods will be used including Principal Component Analysis (PCA), Factor Analysis (FA), Geometric Mean Aggregation and CI techniques such as, Fuzzy c-Means (FCM) and Vector Quantization (VQ). These methods will be contrasted and compared to introduce a valuable tool for weighting and aggregating the quantitative elements of future SCIs, and it would change the norm when ranking and classifying nations.

- To compare and contrast the performance of different missing data imputation methods including two special FCM techniques that is, the Optimal Completion Strategy (OCS), the Nearest Prototype Strategy (NPS). The results are compared against statistical imputation techniques namely;

the Expectation Maximisation (EM), Multiple Imputation (MI), Nearest Neighbour (NN) and Multiple Regression (MR).

- To investigate the performance of different prediction and forecasting methods to assess the most appropriate technique for forecasting KBE competitiveness performance given the limited data sets available. Time Series, Cross Sectional (TSCS) Panel Data, ANN and SOM will be investigated to create a Unified Knowledge Economy Forecast Map *(UKFM)*.

- To introduce a novel macro knowledge capacity building and competitiveness framework by constructing an Intelligent Syntactic Composite Indicators *(iSCI)* for any nation to share their knowledge, monitor their progress, track their KBE competitiveness to improve their overall welfare.

- To simplify and calibrate the final developed model, a robustness analysis will be performed using Monte Carlo simulation, as an appropriate and justifiable model robustness technique.

- To validate the effectiveness of the introduced *iSCI* and *UKCI* frameworks and to evaluate its strengths and weaknesses. Economies in the Middle East and North Africa (MENA) region are used as case study.

In attaining the above goals the current research study makes a contribution to producing a novel and intelligent indicator, would be suitable for any country with different cultural, socio-economic and technical conditions. It is envisaged to assist such economies in establishing and monitoring a suitable and competitive knowledge based economy. This research uses real data sets to illustrate constructing the major components of the proposed index, which includes the qualitative taxonomy, missing values imputations, the weighting, aggregation and forecasting of the suggested *UKCI* variables.

1.4 Research Novelty and Contributions

This work examined many CI techniques before it delivered an innovative decision making tool branded as the Intelligent Synthetic Composite Indicators (*iSCI*). An application of the *iSCI* was put-forward to develop a Unified Macro-Knowledge Competitiveness Indicator *(UKCI)*. The *UKCI* consists of 80 structural and qualitative variables that benchmark how a KBE compares with other countries. The qualitative taxonomy of the UKCI was developed using a novel approach coined as Fuzzy Proximity Knowledge Mining (FPKM). The ranking is undertaken for a group of 73 countries that include almost all of developed economies and the MENA region countries. Furthermore, this work used advance econometrics analysis with CI techniques to create a Unified Knowledge Economy Forecast Map *(UKFM)*. Hence, the main contributions of this thesis are:

- Identification and analysis of the KBE competitiveness on a macro-level.
- Introduced and coined the concept of Fuzzy Proximity Knowledge Mining (FPKM) process to establish an intelligent qualitative taxonomy to build future SCI.
- Introduction of the novel Intelligent Synthetic Composite Indicators *(iSCI)*, with a real case study and validation.
- Using the *(iSCI)* concept, unified several complex, multi-dimensional macro-knowledge indices, into a data-driven, and unbiased KBE indicator, hence the *(UKCI)*
- Investigate and determine frequent and abnormal KBE behaviours. The approach is based on visualising and clustering data sets in a format suitable for classifying and identifying abnormalities.
- Examine different forecasting models to forecast future KBE competitiveness based on Panel Data: Time Series Cross Sectional (TSCS), Multiple Regression, Feed forward ANN model and SOM techniques to produce the *(UKFM)*, This map can be used to visualise, identify or evaluate stable, progressing or accelerating KBEs.

  Overall, the results obtained in this thesis, suggest novel, unbiased and intelligent methods which can be instantly utilised to build future SCIs in many other fields.

1.5 Thesis Outline

This thesis consists of eight chapters that are organised as follows: In Chapter 2, a review of the existing attempts and research studies on the tools and methods used to evaluate, monitor and forecast progress and competitiveness is given. The review presents the existing studies on the importance and difficulties of evaluating KBE competitiveness, representation, recognition, and the techniques used. In addition, the major techniques used to develop progress measures including statistical, computational intelligence and hybrid methods are detailed.

Chapter 3 tracks, examines and reasons the theories of knowledge, macro-knowledge and competitiveness. The focus is on the types of knowledge that is shared between individuals in a certain society and the "tacit" versus "explicit" macro-knowledge, from a theoretical and practical perspective, to establish a working relationship between these important concepts. This reasoning is to serve as a laying foundation to why and how the interest for a unified measure for macro-knowledge competitiveness started to surface as an essential measure for growth and progress.

The approach taking to construct the qualitative taxonomy or the theoretical framework is explained in Chapter 4. This chapter introduces the utilisations of fuzzy proximity knowledge mining, to build the SCI initially to establish significant, consistent and sound indicator. The suggested taxonomy will serve as a basis of mining the net for selecting, inclusion/exclusion and unifications of qualitative variables from various knowledge competitiveness sources.

In Chapter 5 the data selection, acquisitions process, the details of its descriptions and analysis for the study are explained; followed by an overview of some existing techniques which are used in data treatments, weighting aggregations and predictions. The chapter begins by presenting the traditional techniques such as PCA, Panel Data predictor. Then, different CI methods such as FCM, VQ and ANN techniques used in this thesis are put forward with discussions of their benefits of use in missing data imputations, weighting, aggregations and forecasting.

In Chapter 6 covers the main contributions of this thesis where the methods of aggregations are identified and results are presented for three different weighting and aggregation models. The chapter starts with an overview of simple rule based systems using adaptive neuro-fuzzy systems (ANFIS), followed by the approach taken to produce an *iSCI* to identify the centroid of homogeneously clustered nations using FCM. This chapter closes by comparing and validating the robustness of the proposed framework against two statistical models as a case study. The chapter concludes that the proposed framework and the empirical case study for developing future composite indicators was successfully constructed using CI methods to combine the efforts of non-linear, multi-dimensional variables, into a new *UKCI*.

In Chapter 7 the results of the predictive and forecasting models are presented and validated using visual heat maps and radar charts. A comparison and accuracy results between four prediction models are made to find the best model to predict the future progress of any KBE regardless of limited or missing data points. This chapter closes by presenting a Unified Knowledge Economy Forecasting Map (UKFM), using SOM. The proposed forecasting model has the capability of aggregating major KBE indicators into a unified meaningful map that places any KBE in its league. The UKFM reflects the overall position of homogeneous knowledge economies, and it can be used to visualise, identify or evaluate stable, progressing or accelerating KBEs. Finally, the conclusions arise from this thesis and the formulates of some future research directions are presented in Chapter 8.

Chapter 2

Literature Review

2.1 Introduction

This chapter reviews existing attempts and research studies on the tools and methods used to evaluate, monitor and forecast progress and competitiveness. The review presents the existing studies on the importance and difficulties of evaluating KBE competitiveness, representation, recognition, and the techniques used. In addition, the rise and use of SCI for the purpose of measuring progress in nations are reviewed. This chapter is structured as follows; in Section 2.2, the different efforts and the difficulties and faults encountered in such attempts are reviewed. Some literature on indicators classifications, representation, and the levels of groupings with emphasis on the SCI are reviewed in Sections 2.3. In Section 2.4 and Section 2.5 the major Statistical and CI techniques used to develop progress measures are highlighted respectively. The different attempts and methods used to predict and forecast within this domain using both statistical and CI techniques are discussed in Section 2.6. A Summary of this chapter is presented in Section 2.7.

2.2 Evaluating Knowledge Based Economy

Monitoring and evaluating the overall performance for KBE has become very important, since it promotes strategic development and progress. However, many pressing questions arise in light of this. For example, how a nation is doing in its endeavour to become a knowledge economy? How can a nation realise its potentials and capitalise on it? How can nations formulate a decisive plan and tools to establish, evaluate, forecast, monitor and expand its knowledge competitiveness horizons? What are the advantages a nation can gain by monitoring its KBE competitiveness? The answers to the above mentioned, and more similar questions is very important, yet it is not easy to do; as a matter of fact, finding the answers to these questions is the main business or concerns for many leading organisations. The Economic Co-operation and Development (OECD) core practice and mission is to analyse the KBE indicators to understand the dynamics of the KBE and its relationship to traditional economics (Foray and Lundvall, 1996). According to the UN (2010) "there is still currently no internationally agreed on framework for measuring the extent to which an economy or society is a knowledge based". Hence, a framework to bring together the existing models and a research leading and contributing to the debate on this topic is highly needed.

2.3 The Rise of Synthetic Composite Indicators

Indicators are progress measurement tools, usually made to provide a more precise and consistent signal of change for a certain domain than the use of raw data on their own, by summarising information about such domain or subject of interest using statistical measure, e.g. one can measure the level of education or inflation (Statistics.gov.uk, 2010). In general, indicators are classified into two main categories: Individual "simple" and aggregate "synthetic composite" or 2nd generation indicators (Saltelli et al., 2012; Urra, 2007; Arndt and Oman, 2006; M. Saisana and Tarantola, 2005). Synthetic Composite Indicators are assessment tools, usually constructed to evaluate and contrast country performance, by aggregating abstract issues in many areas such as the economy, education, technology and innovation (OECD, 2008b). SCI are developed by using qualitative (qualitative synthetic indicators) or quantitative (actual synthetic indexes) methods. SCI may help detect related essential facts, which may not be measured by the basic grouping of simple indices since this combination does not incorporate the inner arrangement of the structure, nor it does clarify how simple indicators relate to one another (Cumbrera et al., 2008). Furthermore, the European Commission Joint Research Centre (Centre, 2012), has put three levels of indicator groupings as follows: Individual, thematic and composite indicators. Individual indicators represent a menu of separate indicators or statistics. This can be seen as a first step in stockpiling existing quantitative information. The thematic indicators are individual indicators grouped together around a specific area or theme. They are generally presented individually rather than synthesised in a composite. While the composite indicators are formed when thematic indicators are compiled into a synthetic index and presented as a single composite measure (Saltelli et al., 2012).

2.4 SCI Developments Methods

In general the concept of SCI can be viewed as a paradigm of reducing multi-dimensional and non-linear inputs (variables) into a single and meaningful output that can be interpreted by public officials, business leaders, decision makers and ordinary citizens. Even though the issue seems simple, however, *"the methods for aggregating vast amounts of empirical data remain rather crude"* (Cherchye and Kuosmanen, 2004). This aggregated output (hence, the final SCI value) is usually represented and communicated as a single numerical score and/or an ordinal rank. To arrive to this final output value, the index must go through rigorous and monotonous development steps, such as what variables to include/exclude, what to substitute for missing data, how much weight to give to each variable etc. In general, the domain that covers the overall process of understanding the information implied in the variables to measure progress or competitiveness between countries for example, can be investigated along two parallel paths: first, by jointly studying the constituent elements which makes a SCI. Second, by studying the countries in terms of similarity between different elements. (Nardo et al., 2005).

- **The Constituent Elements Similarity Process**: This track usually involves a couple of steps which starts with making sure that the available variables are adequate, enough and well defined to depict and explain the nested composition of the composite indicator appropriate to describe and develop a new index to measure the progress in a certain domain. This is usually accomplished based on "expert opinions" or based on studying the arithmetic structure of the available dataset. Classically, PCA, FA or Cronbach Coefficient Alpha can be used for that purpose.

- **The Distance Measure or Countries Clustering Process**: This track involves clustering countries in terms of similarity between different elements using different clustering techniques. The clustering is usually based on distance measures such as Euclidean, Squared Euclidean or City Block etc. These measures and its detailed rules and methods will be discussed later in Section 5.4.4.

Along these two parallel tracks come the details of the different methods and techniques that can be used to construct a SCI to measure the progress or competitiveness in a certain domain. With the details comes the subjectivity, disagreements and biases of which method and why. Most of the research which has been carried out to deal with constructing SCI is done using statistical or knowledge rule based techniques. Statistical techniques are used to find the dependence and correlations between the variables collected to measure a level of progress, to treat missing values, detect outliers and ultimately to weight and aggregate the composite index. Another common approach for combining multiple attributes is through the use of a set of knowledge rules like IF-THEN to reflect the experts judgement on the input values and a panel of "experts" would set the weight to be assigned to each input. Even though this method was helpful in predicting an output in a small scale, it however offers very little help when the number of inputs goes beyond the human ability to generate so many rules or to reduce the rules to give the best answer; so it is to a certain extent become impractical and a matter of a personal judgement or a best guess intuition.

2.4.1 Multivariate Analysis

Niwa and Tomizawa (1996) constructed the General Indicator of Science and Technology (S&T). Multivariate analysis techniques such as PCA and FA was applied to a set of fourteen related indicators to investigate its structure and extract a single or a small number of indicators. The authors claimed that using multivariate analysis was advantageous for examining and interpreting the characteristics of indicators, and for producing the desired S&T composite index. The mentioned analysis was performed for 5 economies only, the U.S., Japan, Germany, France and the U.K. which is very limited and can't be generalized to extract the needed facts regarding the features of S&T action globally or for the world countries comparisons. Dyba (2000) used FA and Cronbach Coefficient Alpha subsequently, to analyse and produce a measurement tool for the key elements of success in improvement of software process. The measures were found to have satisfactory testing properties. However, a recent study by Grupp and Schubert (2010) negated these findings and claimed that using multivariate analysis for exploratory and weighting purposes may lead to drastically differing ranking results when compared with other methods. Some other disadvantages are that multivariate analysis are sensitive to presence of outliers, small-sample and data modifications such as revisions and updates. Despite these disadvantages, multivariate analysis still used to develop many indices, for instance the ICT Development Index (IDI) created by ITU uses PCA to set the weight for the indicators and sub-indices included in the IDI (ITU, 2012). Furthermore, a collaborative work between, Yale University Centre for Environmental Law and Policy, and the Centre for Earth Information Science Information Network, at Columbia University has produced in year 2000 the first environmental performance composite index called the Environmental Sustainability Index (ESI), which was later replaced by the Environmental Performance Index (EPI) and 'Trend EPI'. Both EPI and 'Trend EPI' are using PCA and FA beside other methods to conduct the exploratory analysis and to help in setting the weighting for these indices (Yale, 2012).

2.4.2 Statistical Weighting Techniques

Use of weighting methods is crucial for developing composite indicators, as proper variable weightings gives a better illustrative of the outcomes ranks or scores. However, debate continues about the best strategies for weighting variables. The following is a critical review of the most widely used statistical weighting techniques:

2.4.2.1 Equal Weighting

Many composite indicators depend on Equal Weighting (EW) or "variance-equal weighting," where all variables are giving equal weight. Babbie (1995) supports such practise and recommend it to become the standard for setting the weights when constructing CSIs. Also Hopkins (1991) tout EW for its simplistic approach and he justify that in accordance with the Occam's razor principle of 'simple is best'. However, in a detailed article Cherchye et al. (2007a) raise many flags against such practise and they argue that, just because we can't obtain agreement on weights, does not constitute using "fundamentally flawed" method. They also argue that EW is out of the core debate in SCIs development and they show how EW interferes with the basics of normalization process. EW allows for perfect exchange among variables, thus disregards the balancing nature of other variables (Cerulli and Filippetti, 2012). For example the Human Development Index (HDI) (Bhanojirao, 1991), and the Innovation Union Scoreboard (IUS) (EuropeanCommission, 2011) are using the equal weighting to arrive to the aggregated HDI and IUS scores. This technique basically denotes that all variables are contributing equally in the aggregation, which could mask the lack of an arithmetic or a practical foundation, e.g. when there is inadequate information of the underlying relationships or absence of agreement on the substitution. However, equal weighting does not imply 'no weights', but indirectly indicates that the variables influences are equal, hence the importance of the underlying variables are the same. This could result in an unbalanced structure in the composite index (OECD, 2008a; Grupp and Schubert, 2010). EW has been widely criticized by many more including Panigrahi and Sivramkrishna (2002); Cherchye et al. (2004, 2007b); Hatefi and Torabi (2010); Belhadj (2012). In conclusion and at most, EW offers the easiest solution, but definitely not the best.

2.4.2.2 Data Envelopment Analysis and Benefit of the Doubt

Data Envelopment Analysis (DEA) uses linear programming to spot the leaders from a set of countries to be used as a benchmark to measure the progress for the rest of countries in a dataset. A few attempts by Cherchye and Kuosmanen (2004); Archibugi et al. (2009) were steered towards that direction using the statistical benchmarking or "benefit-of-the-doubt" weighting technique which is an application of DEA. The benchmarking technique was also used by Mahlberg and Obersteiner (2001) as an alternative method to remeasure the HDI and the Sustainable Development Index (SDI) which was used by Cherchye and Kuosmanen (2004) to identify the benchmark countries. The DEA weighting technique suffers from 'an over performance' problem, because it groups the economy to be predicted with highest neighbour and hence giving it a high weight and or what they call the benefit-of-the-doubt weighting, which eventually gives the selected economy a high score and hence a false 'shiny progress picture'. Other disadvantages of the benchmarking technique is that this method is sensitive to the type of normalizations technique and it depends on the benchmark; if any of these changes, the scores are likely to give different weighting, hence, different country ranking (Nardo et al., 2005).

2.4.2.3 Regression Based Weighting

Porter and Stern (1999) collected and used survey data to measure national innovative capacity. They computed each executive opinion survey as the average reply by respondents for each economy and they used Analysis Of Variance (ANOVA) to assess the dependability of the approach for each survey measure. Regression was used to regress each survey replies on a full set of economy dummy variables, computing the portion of difference between responses that concluded from regular economy-level dissimilarities. This procedure, although appropriate for a large set of variables of diverse natures, assumes that the behaviour is linear and requires focus on the individuality of the independent variables.

2.4.2.4 Unobserved Components Model

The Unobserved Components Models (UCM), is similar to the familiar regression analysis, however, the major dissimilarity is in the response variable, which is anonymous in UCM. Weights with UCM are assigned by assessing the maximum likelihood function of the base indices. The UCM suffers from poor dependability and robustness of results which depends on the disposal of large dataset or long time horizons. It works well with independent sub-indicators, but poor with highly correlated sub-indicators (Nardo et al., 2005).

2.4.3 Participatory Weighting Techniques

In addition to the purely statistical methods mentioned in the previous sections, alternatively, there are participatory methods which considers experts, politicians or citizens opinions to assign weights. This approach is very subjective and depends on the people's beliefs and opinions of how to assign the weights. Below is a brief review of the most famous methods of the participatory weighting techniques:

2.4.3.1 Budget Allocation Process

Also known as Experts Allocation Process (EAP) where experts on a given domain of expertise (e.g. economy, education, corruptions, etc.) are joined together and given a pool of one hundred points or what is known as a "budget" and they are asked to allocate it to the indicator set. This method relies on the experience and subjective judgement of the relative importance of the respective indicators. Weights are calculated as average budgets. BAP is good if the number of indicators is between 10-12. However, if there is many indicators to consider, then this method can cause mental strain to the people who are expected to assign the weights. The Index of Economic Well-Being (Osberg and Sharpe, 2002) used to employ "Expert weighting" BAP, but it has been replaced by equal weighting because of criticisms regarding the weights decided (Sharpe and Andrews, 2012).

2.4.3.2 Public Opinion

Practically, opinion polls concentrate on the idea of "concern" where individuals are requested to state their grade of worry (e.g. small or big) about certain issues. As with expert evaluations, the budget allocation process may well also be utilised in public opinion polls. But it is more challenging to request the public to assign a hundred marks to numerous individual indicators to state a grade of worry concerning a certain problem (Mitchell et al., 1995). A study by Mitchell et al. (1995) used the opinion of seventeen industrialized countries citizens to challenge the World Health Organization (WHO) ranking of 191 countries health systems which was based on the advice of health experts. The study concluded that there is insignificant association between WHO scores and the well-being of the citizens who use these health organisations.

2.4.3.3 Conjoint Analysis

This is a segmentation multivariate data analysis technique grounded on scenarios. A scenario could be a certain set of scores for individual variables. The first choice is then segmented marked and assigned to the assessment. Even though this practice employ statistical analysis to handle the data, it depends on the opinion of consulted citizens, politicians or experts. Each of these individuals is given a different group of variables to evaluate and then to choose their favourites. (Saltelli et al., 2012).

2.4.3.4 Analytic Hierarchy Process

This is a technique used for multi-attribute decision making, and to establish measures for qualitative and quantitative features of a problem which are integrated into the assessment practice. AHP facilitates the decomposition of a problem into a hierarchical structure. Weights are assigned by the importance of a certain domain. Hence, it represent the trade-off across indicators, where an expert may show willingness to forego a given variable in exchange for another because they are not important coefficients. This method rely on people's beliefs; however, beliefs are not always consistent. AHP is based on a redundant process, so it is computationally costly, in addition, redundancy allows for a judgement errors, and an inconsistency ratio (OECD, 2008a). Despite these disadvantages, AHP is still popular, and it has been used by British Airways to choose the entertainment system vendor for its entire fleet of air planes. Also, the Nuclear Regulatory Commission (NRC) of the US used it to allocate a large budget for their computing requirements and information technology projects (Saaty, 2008).

2.4.4 Statistical Aggregation Techniques

The literature of SCI's offers many examples of aggregation techniques. However, one of two major aggregation techniques is usually used: additive (linear) or multiplicative (geometric or non-linear) techniques. The additive aggregation is basically summing up all weighted indicators and sub-indicator to produce a comparison score or ranks between nations. On the other hand multiplicative aggregation is the product raised to the power of the weighted indicators or sub-indicators. Additive aggregation offer full compensability between aggregated indicators, where poor performance indicators can be covered or "compensated for" by the other indicators which have significantly higher value. Such trade-off could thus result in a biased composite indicator. On the other hand, multiplicative aggregation such as the Weighted Geometric Mean (WGM) is less compensatory and touted as superior to the additive method. However, this technique could inflate the overall scores of a SCI, despite a slight improvement in its variables (Munda and Nardo, 2005). The additive methods especially the weighted sum aggregation are usually the preferred choice for transparency, simplicity and ease of interpretations and use even by novices. However, detailed examination of additive and multiplicative methods by Ebert and Welsch (2004) and Zhou et al. (2006) showed that the WGM method often has better properties than the weighted sum method. The Innovation Union Scoreboard (EuropeanCommission, 2011) was aggregated using the WGM.

Another multiplicative aggregation method proposed by Munda and Nardo (2009) is known as Non-compensatory Multi-Criteria Approach (NCMCA) which is trying to make a balance between the cases for a certain objective, e.g. monitoring the level of smog and pollution and at the same time increase the economic well-being of a certain entity. Even though this approach is more advanced, however, it is commonly more complicated to compute and the interpretation of the results is less intuitive (Nussbaumer et al., 2012).

2.5 Computational Intelligence Techniques

As an alternate to the statistical techniques, Computational Intelligence (CI) methods are becoming the trend for their precision in predictions, clustering, modelling and trend analysis; some methods are more popular than others and the proceeding sections presents a review of such used methods.

2.5.1 Artificial Neural Networks

Artificial Neural Networks (ANN) are now considered to be the most popular methods to deal with non-linear and ambiguous cases. For example a recent study by Amin et al. (2009) showed that ANN can act as an aggregator to multi inputs to form a single output. The author applied two ensemble methods; the negative correlation learning and bootstrap aggregating (bagging). Experimental results on a number of real-world benchmark problems showed a substantial performance improvement over other aggregator types. The authors in (Wilson et al., 2002) presented an ANN model, trained using the UK Nationwide House Price Index data to model the projected movements in property prices and to forecast future trends within the housing market. It has been shown that ANN can model any functional linear and non-linear relationship, and that such models are better than regression since regression is essentially a linear technique used to solve non-linear problems. A small scale study by (Sarlin, 2010) used Self Organising Map (SOM) to predict and monitor the financial stability, and sovereign debt for nations. It was concluded that SOM is considered to be a feasible tool for aggregating multiple related variables to visualize and monitor the evolution of economic conditions over time.

2.5.2 Fuzzy Logic

Two related studies carried out by Ammar et al. (2004); Shnaider and Haruvy (2008) suggested that fuzzy logic offers a potential solution to the problems of weighting and aggregating. For example, (Shnaider and Haruvy, 2008) compared the performance of fuzzy logic with linear regression as modelling tools for determining the constituent factors in the assessment of economic growth of national economies. Also, Keller (2008) introduced a multi-input, single-output fuzzy controller to act as an artificial decision maker that operates in a closed-loop system and in real time to forecast and control the dynamics of macroeconomic variables using a fuzzy learning algorithm. It was found that the fuzzy logic yielded more stable and consistent results than those of linear regression.

A few other attempts were aimed at creating fuzzy indicators that use "pure" fuzzy logic to evaluate and rank nations sustainability performance, for example, the sustainability assessment by fuzzy evaluation (Andriantiatsaholiniaina et al., 2004; Phillis et al., 2011) introduced "experts" based fuzzy rules that uses a fuzzy weighted sum of inputs which was computed and assigned to the output variable. Abouelnaga et al. (2010) offered a Nuclear Energy Sustainability Index, using fuzzy logic, which they based it on the same methodology offered by Andriantiatsaholiniaina et al. (2004). Yet again, experts' rule based systems suffer from generic problems such as the subjectivity and possible biases of the experts who devise the rules; the credibility and the hidden identity of who the authors usually call "experts", also the overlapped variables which mostly would generate unmanageable size and tangled rules.

In regards to treating missing data, Hathaway and Bezdek (2001) proposed four new techniques which can be integrated with Fuzzy c-Means (FCM) to allow it to accept and cluster incomplete datasets. Another study by Nuovo (2011) have applied such the aforementioned technique to show that one can indeed use FCM derived strategies, to precisely impute missing data.

2.5.3 Hybrid Techniques

Studies by Kershaw and Rossini (1999) and Dostál (2009) incorporated hybrid methods like fuzzy logic, traditional econometric techniques, and genetic algorithms to develop constant price index and stock market decision machine. The work indicated that such methods could be integrated to present a real alternative to the econometric methods or to improve prediction accuracy. Another integrated methods by Liu (2007) used Multiple-Criteria Decision-Making (MCDM) technique and fuzzy logic to calculate environmental sustainability, rank and cluster nations. The framework considered five components: air, water quality, water quantity, land use and natural resources. Studying such methods, one would expect that the authors had figured a way to extract rules using the MCDM to feed into the fuzzy system to create a coherent and intelligent rules to govern the fuzzy part of the framework. However, the use of MCDM did not serve in creating the rules for the evaluation by the fuzzy logic, instead the framework consisted of two separate techniques that do not complement each other. To set the weights and create the rules the authors resorted to the classic "IF-THEN" fuzzy rule based system which was put by a panel of three "experts" to govern the framework. As stated before, experts rule based systems suffer from common problems such as the partiality and subjectivity of the experts who create the rules. The trustworthiness and the concealed identity of the "experts", in addition to producing large size and entangled rules.

Regularly, the Arithmetic Mean (AM), Geometric Mean (GM), Additive Rules (AR) etc. are used for the purpose of aggregating the variables to form a single value, hence a "composite index" (OECD, 2008a). On other hand, computational intelligence techniques, such as ANN, SOM and fuzzy rule-based systems and recently some hybrid methods such as hesitant fuzzy geometric means, and intuitionistic fuzzy hybrid geometric operators have been proposed and applied to act as an aggregator for multi input-single output systems. Such methods were made to help decision makers to effectively deal with multiple attribute decision making under hesitant or intuitionistic environments (Zhu et al., 2012; Zhao and Wei, 2013).

Panel data regressions or Time-Series Cross-Section (TSCS) regressions mixed with some CI methods has been astonishingly neglected, but a unique study by Pao and Chih (2006) concluded that ANN models can be used to solve panel data regression, and that it would allow to construct and test sophisticated models than purely cross-sectional or time-series data to solve debt policy forecasting problems. Saisana and Munda (2008) proposed using sensitivity analysis when deciding on what to measure, and what to include/exclude from different individual indicators. Herrero et al. (2011) incorporated three different methods, including Cooperative Maximum Likelihood Hebbian Learning, SOM and Curvilinear Component Analysis to select the most appropriate variables to forecast the political risk for most of the world's nations. However, the suggested techniques are either heuristics or too technical which do not measure the reality of the developments pressing issues.

2.6 Prediction and Forecasting Techniques

The literature on forecasting techniques can be grouped into seven main categories:

- Quantitative (e.g. arithmetic average, moving average, and simple exponential smoothing etc.).
- Time series (e.g. trend estimation, AutoRegressive Moving Average (ARMA), AutoRegressive Integrated Moving Average (ARIMA), and growth curve etc.)
- Econometrics (e.g. regression analysis, AutoRegressive Moving Average with exogenous inputs (ARMAX), and Panel Data Analysis etc.).
- Judgemental (e.g. Delphi method, Scenario building, and experts opinion and judgement etc.).
- Naïve (e.g. Naïve Forecast 1 (NF1) and Naïve Forecast 2 (NF2) (Makridakis et al., 1998).
- Computational Intelligence techniques, such as, Support Vector Machines (SVM), Artificial Neural Networks (ANN), fuzzy time series and Adaptive Neuro-Fuzzy Inference System (ANFIS) etc.

ANN methods are becoming very popular forecasting tools for their versatility, and solvency to model temporal and non-linear datasets effectively and accurately (West and Dellana, 2011). Many researchers have introduced various CI techniques and ANN models to forecast with complex, non-linear, short time series or missing data. For example Ilonen et al. (2006) integrated Bass model, ANN and Kohonen SOM to forecast the diffusion of innovations in nations. SOM is used in a variety of forecasting applications, such as short-term electricity load (Yadav and Srinivasan, 2011), countries' political risk or instability (Smith, 2002) and diffusion of innovations in nations (Ilonen et al., 2006). In two recent studies, Sarlin (2010) and Sarlin (2011) use ANN to predict and monitor the financial stability and sovereign debt for nations. They found that SOM model is capable of categorizing macroeconomic time-series data according to vulnerability for an imminent debt crisis. The SOM is considered to be a feasible tool for aggregating multiple related variables to visualize and monitor the evolution of economic conditions over time. However, Herrero et al. (Herrero et al., 2011) argued that SOM results were inconsistent and inferior compared to other techniques such as PCA when they were applied to assess and project the political risk for nations.

It has been established that ANN can also be used to forecast linear and non-linear relationship (Adeodato et al., 2009; Ilonen et al., 2006; Crone et al., 2011). ANN models outperformed several traditional statistical techniques, since these techniques are essentially linear techniques and they require long time-series to be able to predict successfully. However, building an ANN for forecasting with short time periods is not a straight forward task, because many things must be taken into consideration to get the desired accuracy, such as the ANN type, the layers counts, the counts of hidden neurons in each layer, the training method, the activation function, data preparations and divisions etc. (Wilson et al., 2002). Even though the ANN method is ideal for short time-series analysis, the existing applications of the SOM in macro-economic analysis consist of only few papers (Sarlin, 2011). SOM was used in a variety of forecasting applications namely countries political risk or instability (Smith, 2002) and diffusion of innovations in nations (Ilonen et al., 2006).

It has been established that by combining the fuzzy qualitative approach with the neural networks adaptive capabilities through the use of ANFIS model one can produce an accurate prediction and forecasting to many multi-variate and non-linear problems. A recent study by Bektas Ekici and Aksoy (2011) used ANFIS to create a model to forecast building energy consumption in a cold region. It is concluded that ANFIS model was successful predicting the energy need to be addressed in the initial-design stage of buildings to produce energy efficient buildings structure. Cheng et al. (2007) used ANFIS to predict the investors' actions in a stock market in anticipation of event-changes. They concluded that there is a potential of the ANFIS model in financial applications, however there results suggest that sometimes the investor's behaviour are too complex for ANFIS to deal with. Keles et al. (2008) also tuned and trained ANFIS model to forecast the domestic debt for a certain economy to help prevent the collapse of such economy and to give the decision makers an accurate forecasting tool. Even though the above methods are ideal for short time-series, and variables predictions the existing applications of the ANN, or ANFIS in macro-knowledge analysis consist of only few papers (Sarlin, 2011), and to the best of the authors knowledge none applied ANN or ANFIS to the forecasting or prediction and validations of macro-knowledge progress and competitiveness.

2.7 Summary

The knowledge gathered through this literature review suggests that it is possible to use CI techniques to develop future SCI for the purpose of accurate data driven and non-bias KBE progress indicators. Although the use of statistical methods is popular in forming composite indicators and predicting KBE future state, there are some problems associated with their utilization. For instance, statistical models suffer from many issues such as heteroscedasticity, multicollinearity and bias or subjective results. Hence, this study suggests experimenting with CI techniques in comparison with statistical techniques to construct a unified knowledge progress indicator, to answer the growing demand for a "one for all" solution to measure, evaluate and predict the future performance of a KBE.

Chapter 3

Reasoning and Epistemology

3.1 Introduction

Epistemology is the philosophical branch that studies the nature and foundation of knowledge, and tries to define what human knowledge really means (Klein, 2005). There are a number of theories which relates to epistemology of knowledge and aiming at revealing all its domains philosophically. This chapter is an attempt to track, examine and reason the theories of knowledge, macro-knowledge and competitiveness. However, the focus will be - without losing track in an abstract and philosophical debate - on the type of knowledge that is shared between individuals in a certain society and the "tacit" versus "explicit" macro-knowledge, from a practical perspective, to establish a working relationship between these important concepts. This reasoning is to serve as a foundation as to why and how the interest for macro-knowledge started to surface as an essential fuel for economic growth and competitiveness on a macro rather than micro level. However, for specific and detailed drill down on micro- type of knowledge epistemology, such as knowledge by description, knowledge-that, knowledge-(Whether, who, why, what, how etc.), and to track the different attempts to understand whatever is most fundamentally understandable about the nature and availability of such knowledge, the reader is advised to refer to an article written by Hetherington (2012).

The layout of this chapter is organized as follows; in Section 3.2, two distinct schools for theories of individual or micro-knowledge will be distinguished and discussed. In Section 3.3, the theory of macro-knowledge, explicit 'hard' and tacit 'soft' macro-knowledge and their sub theories will be discussed. In Section 3.4, the instrumentalist dynamic, pluralist and Porter competitive advantage for nations theory are introduced as a combined solution to unify knowledge economy and competitiveness theories and to explain how this combined efforts would offer a better framework that would take into account emerging and quickly developing nations. The chapter summaries are provided in Section 3.5.

3.2 Knowledge Theories

There are many general kinds or forms of knowledge streams. But the focus in this research is on two theories of knowledge, which can be distinguished in the literature: First, is primordialism epistemology, which also relates to essentialism. Second, is the pluralist epistemological theory. The first view-originates from theories of ethnicity in social science- sees knowledge as fixed, absolute and universal truth (Stewart, 2007). Followers of this school comprehend knowledge as a tangible and unconnected object and that people may have, or have not. In this discreet and rigid approach, knowledge is disconnected from the seeker of knowledge. Primordialism epistemology branches from the positivist theory, which has been the dominate in the 19th century and still strong today especially in natural sciences. Examples and descriptions of how knowledge is being handled originate from explicit disciplines that produce a tangible output (Stenmark, 2001). The pluralist view on the other side, approaches knowledge as a construct of many social phenomena. According to the theorist of this school knowledge cannot be a universal object or "one size, fits all" and that knowledge is gained when shared between people and in practice (Polanyi, 1962).

Spender (1998), who favour a pluralist epistemology, speaks and justify the existence of relationships and interactions between different types of knowledge which can be represented by different views that can explain such relations. By pursuing a pluralist view, there will be several such knowledge frameworks to acknowledge that no single view is capable of creating the "collective truth". Hence, efforts to tailor a "master framework" that can be applied by all will simply not succeed. It also seems reasonable that different knowledge views are relevant in different circumstances and it is therefore significant and important to apply certain view for specific situation, if deemed useful (Spender, 1998).

Many knowledge types have been proposed so far, for example Polanyi (1967), stated in the "Tacit Dimension" article, *"I shall reconsider human knowledge by starting from the fact that we can know more than we can tell"*. It is apparent that Polanyi was aware that knowledge cannot be always articulated in words and that there is hidden form of knowledge that will not show unless it is practised and this form is what he labelled as "tacit knowledge". In accordance, "tacit knowledge", encompasses a range of theoretical metaphors and sensory information that can be brought together to make sense of something, or help form a new model or theory (Polanyi, 2009). Nonaka (1994), on the other hand draw the line between tacit and explicit knowledge. Choo (1998) building on others theories suggested a third type of knowledge he called "cultural knowledge" which he defined as *"The shared beliefs that shape an organization's purpose and identity, and determine the value and significance of new information and knowledge"*. Spender (1996) splits knowledge into tacit, explicit, individual, and collective. He also argues that knowledge remains divided as a concept between the constructivists, positivists and the pragmatists. For pragmatists, ideas demonstrate their values insofar as they enrich human experience.

3.3 Macro-Knowledge Theory

The knowledge theories and debate mentioned in the earlier section has listed some of the main views on the concept of micro, individual or at large at an organisation level, but we are also interested in the bigger picture or the "macro" concept of knowledge. It is intended to search for theories that govern the creation, progress, diffusions and competitiveness of the collective knowledge between individuals and firms in a nation and between nations.

The rapid telecommunications and technological advancements has transformed the concepts of micro, individual, within limited cities or regional knowledge, to a nation-wide, collective or macro-knowledge. Macro-knowledge is being created, shared and circulated by highly skilled labour force known as "knowledge worker". Drucker (1957) invented the term "knowledge worker" referring to a worker who's capital is his knowledge. "knowledge workers" are focused on lifelong learning, not lifelong employments. The free movement of "knowledge workers" from one place to another has given labour an entirely different perspective to create a new phenomenon known as "global citizens" of the world (Goldman, 1999). The international migration of highly skilled individuals with a broad range of educational and occupational backgrounds such as university students, nurses, information technology (IT) professionals, researchers, business executives, managers, and intra-company transferees or in sum "knowledge workers" move on a temporary basis, while others migrate with an intention to settle permanently in the host country (OECD, 2002). Murray et al. (2012), argues that the migration of large numbers of "knowledge workers" creates the "brain drain" or "human capital flight" adverse effects, which is considered as a national cost.

3.3.1 Knowledge Based Economy

Creating and keeping a high stock of knowledge workers creates the information or Knowledge Society (KSOC) (Mattelart, 2003). KSOC depends on the production, distribution and use of knowledge as the key driver of progress and wealth creation on a micro or regional levels (Quantumiii, 2011). To mark, trace, measure and benefit from the efforts of all the micro levels of knowledge created between different societies, economists adapted the term Knowledge Based Economy (KBE). KBE was first introduced in 1995 as a concept in a general meeting between the members of the Organisation for Economic Co-operation and Development (OECD). The Canadian team introduced the title "the knowledge economy" and they had discussed the "new growth theory" and "innovation" as two major concepts constituting the establishment and progress of a KBE. The New Growth Theory (NGT) (Romer, 1989), stresses that economic growth results from the rising returns associated with the creation of new knowledge. The possibility to grow the economy by increasing knowledge instead of capital or labour to generate opportunities for vast growth (Cortright, 2001).

NGT also focuses on the scarcity of row materials and urges for identifying the inner or endogenous factors for growing a healthy economy; knowledge and technological advancements, innovation, research and quality education, as crucial elements to be distributed as a common culture among the people of a society (Romer, 1991b). In accordance to the NGT theory, these factors were later articulated as 4 pillars, composed of an innovation system, an economic incentive regime, Information and Communication Technology (ICT), infrastructure, and education investments. In line with NGT concept, in 1990, the World Bank (WB) shifted its long and historical stance and policies on education for developing countries from one that positioned higher education as a luxury item, to one that recognized and supported higher education as a key development priority (Robertson, 2009). To acknowledge the importance of higher education, innovation, economic incentive regime and ICT, as major pillars contributing to advancing growth and progress in nations, and in order to facilitate and help countries trying to make the transition to become a knowledge based economy. The WB has adapted these four pillars and developed KBE assessment tool known as: Knowledge Assessment Methodology (KAM). The WB KAM was designed to provide a basic assessment of countries' readiness for the knowledge economy, and identifies sectors or specific areas where policy makers may need to focus more attention or future investments (Chen and Dahlman, 2005) and (Chen and Kee, 2005).

As the idea of a knowledge-based rather than a manufacturing-based economy, more recently began to get a traction, the WB KAM assessment tool served the developed nations become more versatile. However, the KAM has failed in three aspects; first, before 1995, The WB was denying loans for countries wanting to upgrade their research and higher education institutions. Such policy has left many underdeveloped and developing nations with deteriorated higher educational systems. The KAM does not in any way make up for the long term failed policy which counted higher education as a luxury item. Second, the world has progressed way beyond the four major pillars as listed in the WB KAM. Economists are now faced with a new type of economy that is based on special types or advanced digital and nano technologies. Some call it the "digital economy", Internet economy, cyber economy, web economy or simply the new economy (Conceio et al., 2001). Third the KAM has failed to measure the competitiveness of KBEs, as new competitors nations aren't content to remain local because, the digital economy enabled all local economy to be global.

It is been predicted that by year 2020, the E7 group, (Brazil, Russia, India, China, Mexico, Indonesia and Turkey) will comprise a larger share of world GDP than the G7 countries (OxfordEconomics, 2011). These nations are not only producing new potential customers, they are creating new competitions.

Many emerging evidences clearly show that the WB KAM represented by the Knowledge Economy Indicator (KEI), is missing some or not covering the full competitiveness spirit that new forging nations are working towards in their pursuit to become a competitive knowledge based economy. Assuming the imminent value of knowledge to an economy, a proper framework to measure such phenomena is needed to guarantee it is measured and administered properly.

3.3.2 Intellectual Capitals

From the gathered literature, it became clear that there were two ways to build a macro-knowledge economy. One way is through the brain power and knowledge of humans by creating and keeping high stock of knowledge workers, or Human Capital (HC). The second way is resource-based, which concerns with creating healthy economy through a combinations of country resources (tangible assets) and Intellectual Capitals (IC) (Teece, 1986). The term IC, combines the idea of the intellect or brain-power with the economic concept of capital, the saving of entitled benefits so that they can be invested in producing more goods and services. From a theoretical point of view, the term refers to measuring the real value and the total performance of intellectual capital's components is essential for any corporate head who knows how high the stakes have become for corporate survival in the knowledge economy and information age. So, the main point is how an organization can affect the firm's stock price using the leverage of intellect (Brooking, 1996).

IC can include the supportive infrastructure such as hardware, software patent, the company reputation etc; skills and knowledge that a company has developed about how to make its goods or services; knowledge workers or groups of employees whose knowledge is deemed critical to a company's continued success; and its aggregation of documents about how they do things such as processes, customers, research results, and other information that might have value for a competitor that is not common knowledge such as customer and supplier relationships, franchises and licences (Rouse, 2007). Just like HC, the characteristics of IC make it quite difficult to measure, and the increasing importance of research in high-tech fields to develop a KBE, supports the argument and need for adding IC measure that incorporates HC.

3.4 Towards A Unified Macro-Knowledge View

Most progress and competitiveness frameworks for nations conceptualise their existence based on two major social science (national identity) theories; Essentialism and Constructivism. The essentialism believes that attitude and thoughts are fixed, outstanding, and originating internally. Hence, it perceives development or progress as a result of fixed manufacturing plants and related activities in certain or few areas within each country. These areas are where the action is in a country, while the rest of the country is lagging behind. The peoples attitude, behaviour and ego is reflected in the society and therefore change, progress, or development is very rigid and slow, so also the methodologies that constitute a framework to track such progress (Stack, 1986).

Instrumentalists see practical purposes, social conditions and human aspiration as the primary causes of change. Instrumentalists believe that knowledge is fluid. Individuals have multiple thoughts, these thoughts shift according to context, such that we learn, then we adapt. Technology is a tool, largely under human control, that can be used for either positive or negative purposes (Surry and Farquhar, 1997). Instrumentalists/pragmatist view a concept or theory by how effectively it explains and predicts occurrences, as opposed to how accurately it describes objective reality (Khun, 1996). Borrowing from instrumentalists' views on knowledge for individuals and societies, sounds a valid approach for fluid, dynamic, non-linear, chaotic and rapidly changing nations. Furthermore, the pluralist theoretical standpoint or the pluralistic epistemology seems as a suitable model for macro-knowledge diffusion in societies. It can be comprehended as opposite to an authoritarian or oligopolistic society, where the power of know-how is concentrated and resolutions are made by few elites in such society.

Constructivism on the other hand, believes that the world is a series of interrelated theories that perceive change, progress or development as imminent and it is therefore partially dynamic. Also, nations do change forcefully or because "they have to", not because "they want to", and change towards progress would have to go through the discovery mode before society accept change (Chandra, 2010). The constructivist paradigm unlocks great opportunities and embraces a wide range of interconnected concepts which imply that progress is similar to the journey of human which start with discovery to gain the benefits of existence. Because the constructivism foundations are well established, the different aspects of economic, political and social change and performance can be tracked and analysed in different ways using advanced econometrics models such as: game theory, formal rational perspectives, case studies or even by using computer generated agent-based models (Lustick, 2000).

It can be concluded that so far, two major paradigms exist for the foundations of progress and other social phenomena's; the "fixed" or "rigid" primordialism or thick and partially fluid constructivism. Existing theories on the measure of hard or explicit outcomes such as infrastructure, economic and political outcomes are mostly driven by the primordialist assumption that these areas are fixed. However, some research started advocating constructivism as a better foundation to guide building future progress and development frameworks (Natoli and Zuhair, 2010). Moreover, most existing theories fell short when it comes to the classifications and measurement of the real forces behind the fast progress in emerging or developing KBEs.

The adoption of a constructivist perspective for measuring the explicit or hard theme of a KBE such as the economy, education, labour market, and public administration is effective because it attempts to incorporate both cooperative action and exclusive knowledge. The explicit and tacit themes that this research study asserts reflect development and knowledge as a variety of factors and collective methods, to hypothesis of an expressive and close to reality illustration of macro-knowledge formation, diffusion and measurement.

Finally, when it comes to the measurement of the intangible assets of nations this research is adapting the famous Porter's national competitive advantage diamond model (Porter, 1998). The diamond model surfaced from a rising agreement that in order to capture the full picture of progress and development for nations, intellectual capital, social capital, innovation, and information technology must be taking into account. The assessment and the methods of assessments devised to measure these soft or tacit aspects will contribute to the advancement and progress of a KBE.

3.4.1 Tacit Macro-Knowledge for Advanced KBE

Macro-knowledge progress in the world is fluid, dynamic, non-linear, chaotic and rapidly changing, and so are nations such as Korea, Brazil, Turkey, China and India. Such nations rely on the "soft" or tacit aspects of their macro-knowledge economy to derive their progress and success, which relies on these nations' huge stock of skilled and "very affordable" knowledge workers. Because soft knowledge is becoming expensive to acquire in developed nations. Such nation rely more on the "hard", tangible or explicit aspect of their macro-knowledge abilities to stay ahead. Also, the high soft skills, high productivity nature of a very competitive KBE, has tipped the scale in favour of other countries with limited land or large labour force, such as the Scandinavian countries, Hong Kong, United Arab Emirates, and Singapore. These small but effective countries managed to forge ahead in competing with huge economies, despite the fact that they lack the natural and human capitals.

One can think of the tacit assets of emerging nations as the individual intangible skills, such as know-how, languages, team work, connection and relations. Such "soft" skills are what distinguish successful individuals, and it is also the case for emerging or developing nations but on a larger and different parameters. Such tacit capabilities are the reservoir or the invisible ladder that such nations are using to climb up in the ranking of competitiveness and development. But what are these intangible, tacit or "soft" components and can we measure them? Another important question is what are the characteristics that distinguish such quickly developing nations? And can we devise a framework that encompasses the "soft" and the "hard" to accurately measure KBE progress?

Given the fact that in any economy physical resources are scarce, the idea of tapping into a country's "soft" resources such as technology, innovation and intellectual capital is such a powerful impact in an economy progress and competitiveness. Hence, an appropriate competitiveness measure needs to account for most aspects of KBE progress so it can serve as a basis for decisions to improve resource allocation. Therefore, anybody embracing a review of competitiveness measurement should not look for what the measure examines, but more importantly, what it falls short of explaining. It is this need- which drives changes in measurement - to determine possible missing elements or ignored modern measurement methods, which could be important for long-term growth, and creation, into a unified and intelligent KBE competitiveness framework.

3.4.2 Competitiveness and Unified KBE Measure

Competitiveness refers to the ability of a firm, industry, region, nation, and regions to contribute to wealth creation and maximization of welfare by selling and supplying goods and services in a given market, while being and remaining exposed to the international competition (Charles, 2013). The mounting sophistication of the global arena and economic difficulties owed to increasing competition between organizations globally has sparked advanced thoughts on how to define and how to measure competitiveness. Decision makers, business people and researchers recognise that advancing competitiveness offers businesses of all kind and hence nations a strategic position for creating long-term success, boost economic progress and the well-being of their citizens. Thus, the incorporations of macro-knowledge and competitiveness under both the Porter's model and the pluralist view imply that the framework should deliver more insights on the overall effect of KBE competitiveness, in addition and contrast to the single constituent elements of KBE progress. Hence, this approach would reduce the drawbacks associated with use of the individual value approach. This research is calling for in addition to unifying the existing indicator as a first step towards understanding how KBE and competitiveness should merge together. The second natural step would be to reverse engineer the existing KBE and competitiveness indicators by taking these existing indicators apart and to rebuild them after doing the necessary test, analysis and modifications and enhancements. This will serve two purposes; to investigate if computational intelligence methods can be used to produce valid and robust future KBE competitiveness indicator, secondly, to re-map the inner components of the under focus indicators to produce what envisaged as a data-driven, objective and robust unified KBE competitiveness index.

Nevertheless, one should bear in mind that "composite indicators" are much like mathematical or computational models. As such, their construction owes more to the craftsmanship of the builder than to universally accepted scientific rules for encoding. With regard to models, the justification for a composite indicator lies in its fitness for the intended purpose, ease of use and wide acceptance" (Rosen, 2005). Therefore, this research is a step towards introducing CI methods for the development of future Synthetic Composite Indicators, in pursuit to measuring an arbitrary concept such as knowledge, competitiveness and development.

To illustrate the benchmark for combining related composite indicators to form the suggested Unified Knowledge Competitiveness Indicator (*UKCI*). This benchmark is parallel to the factor capabilities necessary to clarify the meaning of KBE and competitiveness at the country level within the context of the famous Porter's national competitive advantage model and the pragmatism view. Hence, the qualitative taxonomy which will follow in the next chapter, will be derived by a computational intelligence knowledge mining models that incorporate, unify in accordance to the pluralist theory, and learn to propose a data-driven and an "intelligent" KBE framework.

3.5 Summary

This chapter has reasoned for the use of a more comprehensive definition for KBE to capture not only a country's preparedness for KE but also its macro-knowledge competitiveness on the global landscape. It also argued for an interdisciplinary theories and for the inclusions of Intellectual Capital in addition to the ICT and Innovation to form the "tacit" aspects of KBE competitiveness. This formation was based on the current inadequacies in measurements. As the last two chapters have established, most current views-especially those employing a "hard" commodity or market approach-either downplay or neglect the connections that occur in the real world.

In the next chapter the present research will construct a qualitative taxonomy that adopts a pragmatic approach, which enfolds methodological instrumentalist while confirming with Porter's macro competitive advantage model and evaluation to enrich the theoretical framework. What follows, will deal with issues of overlapping and interconnected concepts. From a KBE development measurement perspective, this approach views KBE competitiveness creation and advancement arising from the ontological relationship of a wide range of factors. These factors will be discovered and utilised through the application of intelligent knowledge mining techniques to guide the development of a robust qualitative taxonomy. The steps and constructs of an intelligent qualitative taxonomy will be discussed and justified in the next chapter.

Chapter 4

Qualitative Taxonomy

4.1 Introduction

The preceding chapters considered different theories and views that have been backing up frameworks to measure developmental and progressive issues such as competitiveness, human and intellectual capital. This research claims that the difference in existing frameworks and measurement techniques, stems from different epistemological views. Also highlighted are the drawbacks associated with using single or static method for the measurement of KBE competitiveness, which creates poor explanatory power and weakened analysis. These and other insufficiencies, have led this research study to adapt an interdisciplinary epistemological structure, where strong evidence has emerged from the gathered literature which calls for a dynamic and unified approach to forecast a KBE capability. The argument is that the constructivist approach would be suitable to explain the dynamic behaviour of vibrant and strongly emerging nations, because the constructivist perceive the world as a series of interrelated and dynamic events, therefore change and progress is a must. The Porter's competitive advantage theory is the method of choice when it comes to measuring competitiveness of KBEs. Finally, the pluralist approach seems feasible for the unification of existed methodologies, and the development of Synthetic Composite Indicators (SCIs). SCI will serve as the tool that can be used for the purpose of measuring KBE competitiveness in a nation and between nations.

The development of a new SCI happens in two major stages; building the qualitative taxonomy, then dealing with the quantitative part. The focus of this chapter will be on devising a novel and intelligent qualitative taxonomy for a new SCI. The suggested taxonomy will serve as a basis for inclusion/exclusion and unifications of empirical variables from various predefined and reliable sources. Later on, specifically in Chapter 6 the data will be collected in accordance with this chapter qualitative taxonomy and framework results. The remaining of this chapter will be as follows; in Section 4.2 will highlight the controversy surrounding SCIs and the shortcoming of existing SCIs for measuring KBE growth and it conclude with the general pros and cons of SCIs. The existing views for establishing a new qualitative taxonomy is covered in Section 4.3. In Section 4.4, the use of CI techniques to build the quantitative side of a new SCI is introduced. It includes the use of Fuzzy Proximity Knowledge Mining (FPKM) methodology for the purpose of devising a non-biased, novel and intelligent method to create a new SCI taxonomy. Section 4.5 illustrates through an empirical case study the use of FPKM to develop the conceptual framework for a new ICT index. In Section 4.6, the FPKM method is generalised to develop the suggested *UKCI* framework and concept map. Summary of this chapter is presented in Section 4.7.

4.2 The Controversy of Synthetic Composite Indicators

Many articles have been published suggesting a new SCI, however, criticisms have been aired regarding these composite indicators' methods of construction and use. While building a composite indicator is not a simple procedure and it has seen a wide variety of "unreasonable performance" such as negative values, uninterpretable solutions and ill-conditioned matrices (Freudenberg, 2003). Preliminary investigation and many concerns raised by Archibugi et al. (2009) suggests that there are four major shortcomings associated with the composite indicators measuring KBE. First, there is little information available on how these indicators relate and interact with each other. Second, the evaluation of the competitiveness of a KBE based on selected indicators is often subjective, as performance expectations (regarded as a measure of competitiveness) are not easily quantifiable. Third, some sources of information contributing to the formation of some of the indicators are open to interpretations (e.g. social-economic factors). Furthermore, most of these indicators report past performance, and they do not predict where a certain KBE is heading giving all known elements. The combination of covering a certain aspect of a KBE, different sub indices measuring different constructs, apparent uncertainties in the imputation of missing data. Furthermore, the lack of clear functional relationships between the indicators and their inability to forecast the future progress has severely limited the applications of statistical techniques. To minimise any potential confusion to decision makers, a new approach is needed, with a capability to incorporate new information to foresee into the future, predict the promising or emerging KBE and to anticipate the declining ones. This will help to ensure a consistent application of judgement and better and wider coverage evaluation of performance, when assessing a nation's potentials and competitiveness using the published indicators.

Pros and Cons of SCI

Pros (Freudenberg, 2003; M. Saisana and Tarantola, 2005; OECD, 2008a)

- SCI can be used to summarise complex or multi-dimensional issues, in view of supporting decision-makers.
- SCI provide the big picture. They can be easier to interpret than trying to find a trend in many separate indicators. They facilitate the task of ranking countries on complex issues.
- Can assess progress of countries over time.
- SCI can assist in attracting potential foreign investors, donors, loans and grants.
- SCI can help attract public interest by providing a summary figure with which to compare the performance across countries and their progress over time.

- SCI could help to reduce the size of a list of indicators or to include more information within the existing size limit. Enable users to compare complex dimensions effectively.

Cons (Freudenberg, 2003; M. Saisana and Tarantola, 2005; Tarantola et al., 2006; Trebilcock and Prado, 2011; OECD, 2008a)

- SCI may lead to inappropriate policies if dimensions of performance that are difficult to measure are ignored and or could give confusing or weak strategy messages if they are inadequately formed or misunderstood.

- The simple big picture results which SCI show, could possibly encourage decision makers to implement naive policy decisions. SCI should be used in along its sub-indicators to draw sophisticated policy conclusions.

- The construction of SCI involves stages where robust and sound methods has to be used for the selection of sub-indicators, choice of taxonomy or theoretical framework, treatment of missing values, weighting and aggregating indicators etc.

- There could be more scope for disagreement among countries about SCI than on individual indices. The selection of constituent elements, weights and imputation of missing values could be the target of political challenge.

- May be misused, e.g. to support a desired policy, if the construction process is not transparent and/or lacks sound conceptual principles.

- The SCI increase the quantity of data needed because data are required for all the sub-indicators and for conducting statistically significant analysis.

In this section the main views of both for and against the use of SCI is reported. The purpose was to show that the debates concerning the development and use of SCI, vary from the power of a SCI to catch peoples attention, to the possible danger of using such measures to make "dull" nations look "smart" or a "troubled" economy looks "healthy". This could happen, due to the fact that the majority of existing SCIs are dependent on, and is very receptive to the imposed subjectivity of the "experts" who devise them. Nevertheless, bad craftsmanship of an automobile does not make it a bad invention! Conversely, SCIs remain a powerful decision making tool, for their ability to bring together many abstract and non-linear issues, and present it as a single figure that can track and show progress in many domains. The determination to use SCI as a tool to measure “KBE competitiveness” means that the focus of the remaining of this thesis, will be on testing alternative SCI development techniques, for the purpose of overcoming the subjectivity, complexion and other pitfalls mentioned earlier. For establishing a simple yet intelligent SCI with forecasting capabilities, this study will employ freely and readily available knowledge from pre-defined sources to make “meaning-driven” learning conceptual framework for a new breed of indicators that would be built using CI techniques. It is envisaged that the proposed SCIs would carry special character as they are “data-driven” and therefore such indicators would have the capabilities to accurately measure the progress in a certain domain and rank nations based on non-bias “learning”. The new SCIs will launch the 3rd generation of composite indicators, which will refer to as the “Intelligent Synthetic Composite Indicators or *iSCI* hereinafter. The main benefit of *iSCI* qualitative taxonomy, is it can be used as the underlying theoretical framework for the proposed Unified Knowledge Competitiveness Indicator (*UKCI*). For this to happen, the SCI must be easy to construct and practical enough to be applied in real scenarios using real variables and datasets related to both macro-knowledge and competitiveness in nations. The following section lists more of the reported pros and cons of SCI controversy.

4.3 Establishing the Qualitative Taxonomy

With controversial concepts such as knowledge, macro-knowledge and competitiveness, it is natural for a number of diverse views and methods to exist. Furthermore, in accordance with the pragmatist approach on such a debatable concepts, what is really important here is not what the theory says, but whether it is deemed practical, useful and fit for the purpose. In this basis, this study asserts that an alternative KBE competitiveness framework does not largely have to justify what lie behind the primary causes and effects. It would rather be a combination of several methods to be unified in order to offer a solution or to help recognise other options we thought not possible in the domain of KBE competitiveness measures. Considering the complexity of the problem, this research implements an interdisciplinary approach when developing the suggested *iSCI*, that is relying on the pluralist and the prgamatists views which needs to be reiterated. Instrumentalists or prgamatists, view a concept or theory by how effectively it explains and predicts occurrences, as opposed to how accurately it describes objective reality (Khun, 1996). Hence, it can be infer that theories should be treated like a "black box" where observed data goes as input, and through which apparent predictions gets produced (Sandhusen, 2008).

Also this thesis rejects theories that treat realistic life disciplines such as economy, education, innovation etc. as static and separate events. Alternatively, the fluid and dynamic characteristics of all aspects in society are settled by their continuous cooperation with each other. Consequently, in keeping with the present research's qualitative taxonomy requirement, it is advantageous to confer two major principles: First, the use of a pluralist approach which encourages the use of SCI under a thorough method, means that the taxonomy must be able to deliver insights on the overall impact of KBE competitiveness as well as on the single components of KBE. By doing so, the weaknesses associated with the single or static value approach can be avoided. Second, many if not all developers of composite indicators relax the fact that these indices should represent and measure realistic events, and a set of synthetic aggregated indicators is not the reality, but it is basically an informative model of it (Saisana and Munda, 2008). It is therefore crucial to use a discipline that can construct better model of reality. One of the major insights of soft computing methods such as fuzzy logic is that many concepts are better defined by words than by mathematics, and fuzzy logic and its graded membership provide a discipline that can construct better model of reality (Cherchye and Kuosmanen, 2004).

4.4 Intelligent Taxonomy Methodology

In the following sections more details about the method of development to construct the proposed qualitative taxonomy for *iSCI* are described.

4.4.1 Fuzzy Proximity Knowledge Mining

In the literature, there are many fully automated methods to create taxonomies and ontology inference to extract knowledge. For example the authors in Drumond and Girardi (2010) suggested using a Hidden Markov model to extract certain knowledge from text to build taxonomical concept hierarchies. Dakka and Ipeirotis (2008) devised an automatic extraction of useful knowledge from databases containing multiple text based documents. In addition some semi-automatic models have been proposed to build an architecture of ontology through "learning" by utilising a company Intranet resources (Kietz et al., 2000).

The process of creating a qualitative taxonomy or theoretical framework for a new indicator is normally triggered by the need for a new index to measure the change in a certain domain. This process is usually developed based on expert opinions or certain stakeholders. However, the aim is to mine useful knowledge from pre-defined sources to make the conceptual framework for a new indicator to fit for a purpose. To achieve this end, it is suggested to use Fuzzy Proximity Knowledge Mining (FPKM) to establish the suggested taxonomy. The suggested FPKM consists of two major steps: focused web mining and fuzzy text matching, which are explained below.

4.4.1.1 Soft Focused Web Mining

This step involves a search for documents and reports that contains variables to measure a certain domain. A web crawler can be utilised to search the hypertext in the web for a certain keyword(s). The crawler usually starts the search from a certain page called a seed URL, and then classifies related documents based on a hierarchical tree with node(s) or page(s). Regular crawlers are inefficient as they could pick large number of copies of the same page if it exist on multiple sites or irrelevant ancestor pages. To address this issue the Soft Focused Crawler (SFC) technique is used[1] (Huberman et al., 1998). The SFC collects web pages that relate to a certain concept represented by a seed URL, d, and then it classifies a page, s, to be retained by calculating the probability of its relevancy or goodness

[1] Statistica 10 - Data Mining software was used to conduct the SFC web mining.

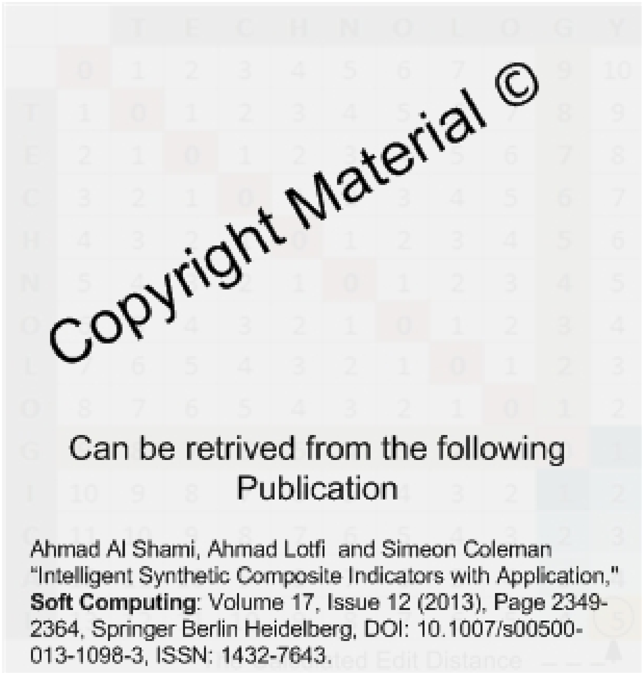


Figure 4.1: Dynamic programming matrix to match two strings.

to the seed URL as:

$$R(d) = \sum_{good(s)} P(s|d) \tag{4.1}$$

the objective is to classify a document to the parent page in the hierarchy with the highest probability $P(s|d)$ (Dunham, 2006).

4.4.1.2 Fuzzy Proximity Text Search and Match

To find and extract the needed variables to be included in the making of the desired index, the crawler results needs to be fetched. The problem can be viewed as a paradigm of "search and match" for a key lexical phrases for certain concept via keywords of the crawler collected sources. This can be achieved by using string searching and matching algorithms.[1]

[1]The rest of this text is censored as a Copyright Material which can be retrieved from the following article: Ahmad Al Shami, Ahmad Lotfi and Simeon Coleman "Intelligent Synthetic Composite Indicators with Application," Soft Computing: Volume 17, Issue 12(2013), Page 2349-2364, Springer Berlin Heidelberg, DOI: 10.1007/s00500-013-1098-3, ISSN: 1432-7643.

An example is presented to clarify the process.[1]

(4.2)

4.5 Developing the Qualitative Taxonomy

The process of creating a qualitative taxonomy or theoretical framework for a new composite indicator is usually triggered by the need for a new index to measure the progress in a certain domain. This process is developed based on experts opinions or certain stakeholder demand. The goal is to mine useful knowledge from predefined sources to make "meaning-driven" learning conceptual framework for a new indicator to measure the progress in a certain domain, without any "experts" interference or biases. To illustrate the application of mining useful knowledge to make "meaning-driven" qualitative taxonomy for a new indicator to measure the progress in a certain domain, without any "experts" interference or biases. Let us assume there is a pressing need for a new index that would unify the efforts between already existing individual indicators to measure the level of ICT and e-services progress between nations. Figure 4.2 depicts a schematic diagram of the steps taken using the suggested methodology of knowledge mining to create a qualitative taxonomy for a unified ICT index. The steps taken can be explained as follows:

1. Web Mining: A crawler software is used and set seed URLs from some predefined web sources. The ITU, WEF and WB web sites are used. The crawler is tuned to search for and export documents containing certain keywords such as"technology, ICT, e-government, communication, networking etc." Unrelated pages are filtered out using SFC on the targeted sources as explained in Section 4.4.1.1.

2. Indicators Analysis: The set of obtained documents resulted from the first step, which contains many file formats (txt, doc, pdf, html etc.) are fed

[1] The rest of this text is censored as a Copyright Material which can be retrieved from the following article: Ahmad Al Shami, Ahmad Lotfi and Simeon Coleman "Intelligent Synthetic Composite Indicators with Application," Soft Computing: Volume 17, Issue 12(2013), Page 2349-2364, Springer Berlin Heidelberg, DOI: 10.1007/s00500-013-1098-3, ISSN: 1432-7643.

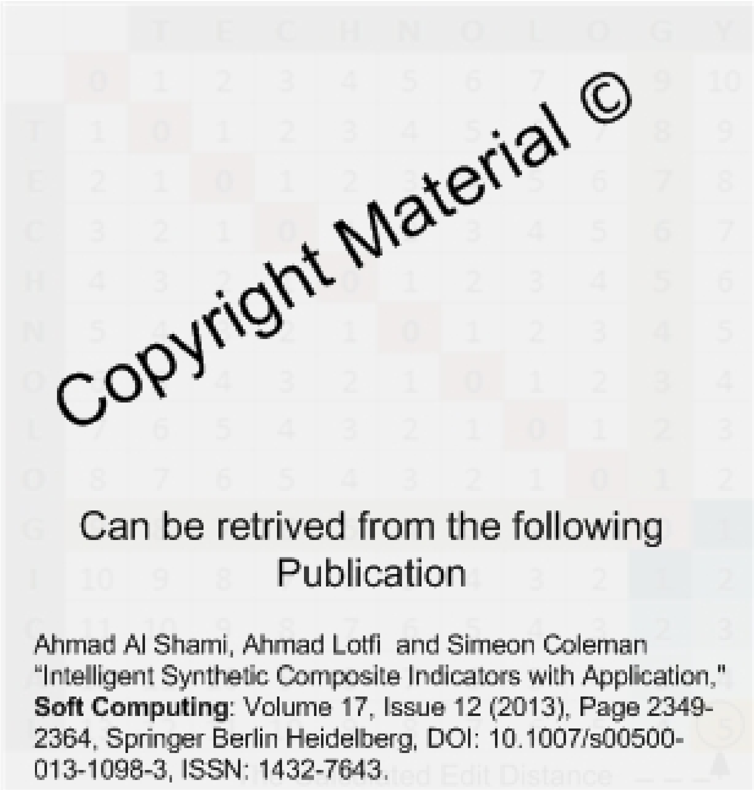


Figure 4.2: Schematic diagram of creating qualitative taxonomy for unified ICT index.

to a Computer Assisted Qualitative Data Analysis software [1] for further analysis.

3. Match, Cluster and Select: In this step further analysis is conducted to find the most common or repeated words between all the obtained indicators. Such analysis includes finding the most common or repeated words, and to cluster the result for the top 10 repeated words, or top 20 ...etc. Also in this step the clustering results are inspected to select a main node or (keyword) for the subject to build an indicator for. In this case the word "technology" is selected as the main subject for the new indicator.

4. Fuzzy Proximity Strings Matching: Finally, the fuzzy text matching techniques as described in Section 4.4.1.2 is utilised to measure the distance,

[1] NVivo10: http://www.qsrinternational.com/products_nvivo.aspx

hence 'match or mismatch' of the selected target word "technology". The result is then listed and zoomed into so the different contexts and the sibling branch of the word "technology" related synonyms and specialisation is alphabetically presented, inspected and grouped. This grouping would represent the "technology" needed variables to form the new indicator qualitative taxonomy.

4.5.1 Variable Sources

As a result of conducting the above steps and as illustrated in Figure 4.2, a handful of Technology, ICT, Networking, and e-services related variables have been collected, but to keep it simple the choices are narrowed to only six sources based on their international presence, significance and reliability. These sources are:

- ITU-IDI: ICT Development Index, developed by the International Telecommunication Union (ITU, 2012).
- WEF-NRI: The World Economic Forum's Networked Readiness Index, result of a collaboration between the World Economic Forum and INSEAD Business School (Dutta and Bilbao-Osori, 2012).
- WB-KEI: Knowledge Economy Index, developed by the World Bank (WB, 2010).
- INS-GII: Global Innovation Index from INSEAD Business School (INSEAD, 2011).
- WEF-GCI: Global Competitiveness Index, from the World Economic Forum (WEF, 2011).
- IMD-WCY: World Competitiveness Yearbook, from the International Institute for Management Development (IMD, 2011).

The constituent elements or components of the above sources indicators are grouped together as pillars and sub-pillars consisting of"hard" and "soft" variables to make its final composite score. The IMD-WCY score for example, is the

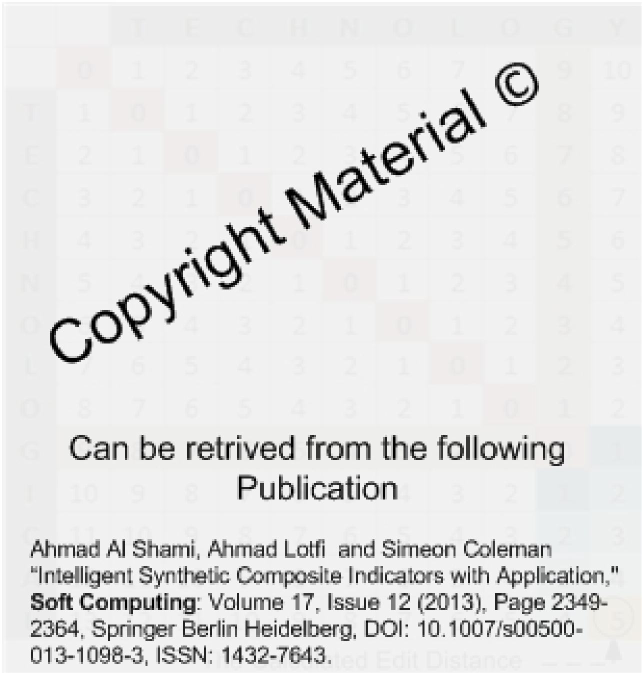


Figure 4.3: Top 10 matched words.

result of more than 300 different aggregated variables, therefore it is challenging to spot and manually match its variables by its name with the rest of the other sources. To solve this issue it is suggested to cluster the constituent elements of these composite indicators based on their word similarity.

4.5.2 Exact Text Matching

To find the degree of similarity by major words or subjects nodes first searched and then listed the "top 10" matched words. Figure 4.3 is the result for the top 10 similar words between all the collected sources which at this instance show that these sources do not include any technology or ICT variables. Nevertheless, a search for the top 20 matched words, revealed and listed the word "technology" between the top 20 matched words as illustrated in Figure 4.4. If however, not successful at finding any related words, we could either dig further to the next top 30 or 40 matched words or we can drop the variables sources and search again for new more technology or ICT related sources.

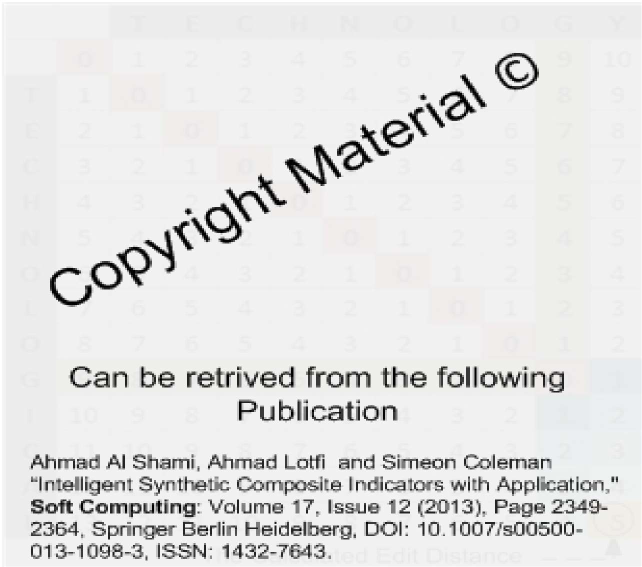


Figure 4.4: Top 20 matched words.

4.5.3 Fuzzy Text Matching

To ensure picking all the technology related and desirable variables to be included in the new ICT index, this study utilised the fuzzy strings proximity matching as described in Section 4.4.1 to find and visually present all possible variables that have any degree of membership with the target word "technology"[1] Figure 4.5 shows the tree map visualisation which lists the different contexts that the targeted word appeared in. Three sibling branch drill-down of the word "technology" related variables are listed, where ICT depicted as the second top term after the targeted word with major branches such as ICT access, price, use etc. With this detailed tree map, one can drill down on a certain word to go to the original source of where such word has appeared and the coverage percentage for any of the target word branches, leaves, its derivatives stems, branches or leaves.

In conclusion the fuzzy knowledge mining model has successfully detected 11 major indicators or sub pillars related to ICT and e-services within the main retained sources. Table 4.1 lists the variables and its sources. Finally, the 11 filtered variables are grouped into one ICT basket which could form different sub-

[1] The WordNet thesaurus built in NVivo 10 software is utilised to search and include the "synonyms" and "specialisation" of the used keyword.

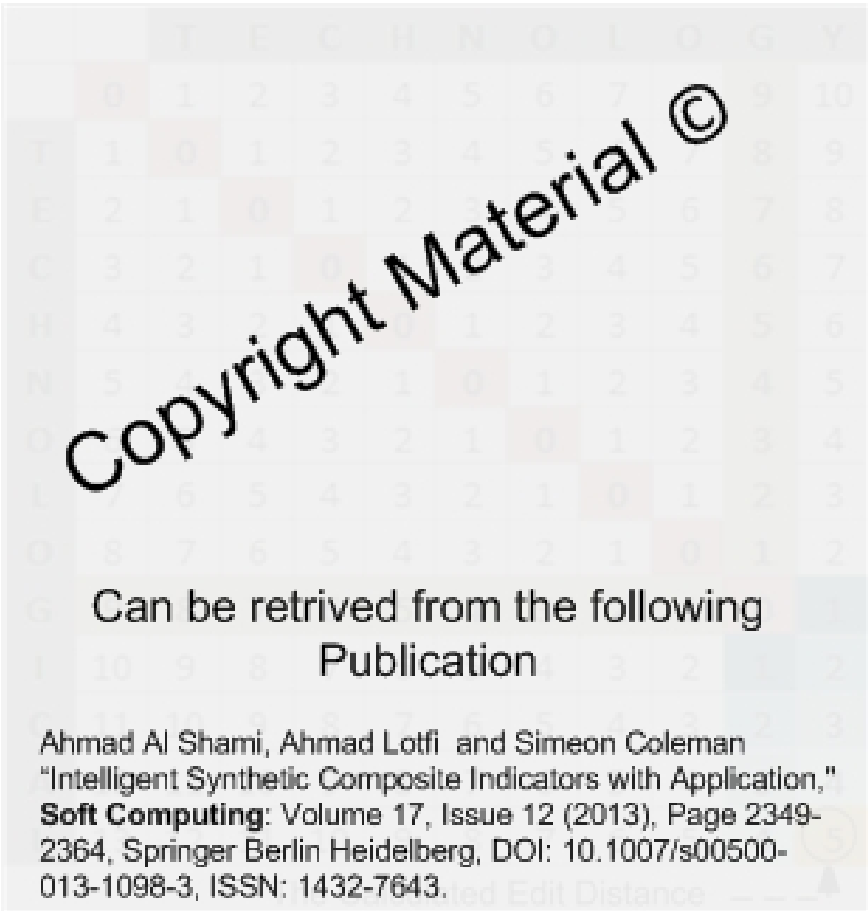


Figure 4.5: Word tree of the "technology" and its derivatives using fuzzy knowledge mining system.

baskets to measure ICT sub-related issues such as ICT environment, readiness, access, use, price and e-services in a country. The basket, and its sub-baskets would represent the qualitative taxonomy for the suggested ICT index as a case study demonstrating the first stage to construct the suggested *iSCI* synthetic composite indicator.

The above process is repeated till all similar variables are spotted, matched and re-mapped to create the suggested *iSCI* "baskets" and "sub-baskets". A basket, and its sub-baskets would represent one of the main pillars of the suggested qualitative taxonomy, demonstrating the first stage to acquire the building blocks to construct the suggested unified KBE competitiveness indicator.

4.5.4 Reasoning and Grouping Matched Indicators

The norm practice for grouping similar variables to create one of the pillars of a certain composite indicator is usually based on ordinal weights which depends on mutually agreed categories (Cherchye et al., 2007a). However, it is shown earlier in this thesis that it is possible to use knowledge mining techniques to automatically reveal a certain or desired qualitative taxonomy i.e. a unified index for technology related variables. The devised and used methods as described in the earlier sections are geared towards mining and learning to help extracting rules from the existing knowledge and competitiveness related composite indicators. The suggested process helped to rectify and propose a new unified qualitative taxonomy, in a completely computationally intelligent way, in order to avoid or minimise human errors or bias interventions.

A quick glance on the technology related tree diagram as depicted in Figure 4.5, one can clearly see that several indicators can be sub-grouped under more detailed headings. For instance: it looks feasible to say that a few keywords related to technology are "ICT", readiness, environment, price, and so on. To be more specific, it is feasible to say that the 11 indicators as listed and coded in Table 4.1 can be re-grouped and based on the same concept as above to form the following five sub-categories:

1. ICT Environment: code B, D, E and F,
2. ICT Readiness: code A and G,

Table 4.1: Technology and ICT filtered variables.

Code	Variable Name	Source	Years	Countries	Scale
A	Technology Readiness	WEF-GCI	2002-2012	159	1-7
B	Infrastructure Environment	WEF-NRI	2003-2012	142	1-7
C	ICT Pillar	WB-KEI	1995-2012	146	0-10
D	ICT Access	ITU-IDI	2002-2011	159	0-10
E	Technology Infrastructure	IMD-WCY	1989-2012	59	0-100
F	ICT Sub-Pillar	INS-GII	2009-2012	125	0-100
G	Gov't Readiness	WEF-NRI	2003-2012	142	1-7
H	ICT Price	ITU-IDI	2002-2011	159	0-10
I	Individual Usage	WEF-NRI	2003-2012	142	1-7
J	Gov't Usage	WEF-NRI	2003-2012	142	1-7
K	ICT Use	ITU-IDI	2002-2011	159	0-10

3. ICT Use: code I and K,
4. ICT Price: code H,
5. E-Services: code C and J.

4.6 Assembling the Unified Framework

The process of extracting, grouping, and sub-grouping are followed as described in the previous sections. This allow to extract the focused objectives of the mined indicators to remap and restructure accordingly.

The process of mining, learning and matching of the exact and fuzzy strings through the devised FPKM model are repeated on each of the main matched words to represent one of the focused objectives of the new unified indicator. Each objective is labelled as a “Basket” i.e. Economy Basket, Education Basket, ICT and e-Services Basket etc. Through a close and cross examination, each basket is then divided to one or more sub-baskets which consist of the indicators of the emerged baskets. The final emerged unified framework consist of 8 main baskets, that are devised into a total of 22 sub-baskets. The baskets of the “soft” vs. the “hard” aspects of the knowledge based economy are then merged to make 2 major themes, which encompass a nation’s “explicit” and “tacit” macro-knowledge. The themes, baskets, sub-baskets indicators and variables would make the major units of the suggested macro-knowledge framework, to be branded as the Unified Macro-Knowledge Competitiveness Indicator (*UKCI*). Table 4.2 explain the hierarchy and the purpose of each level of the suggested *UKCI* framework.

The basic justification of the grouping-inclusion-exclusions and remapping of variables to form such hierarchy for the purpose of building the suggested UKCI framework has already been justified in the previous chapter and within the literature of the already existing composite indicators. For example one can refer to the New Growth Theory (Romer, 1989) to appreciate or validate the inclusion of major baskets such as the technological advancements, innovation, quality education, and economic incentive regime, as crucial elements to be distributed as a common culture among the people to build a healthy knowledge society (Romer, 1991b). A similar approach has also been articulated and adapted by

Table 4.2: The *UKCI* framework development units.

Code	Unit	Rational
1	**UKCI**	The Unified Macro-Knowledge Competitiveness Indicator
2	**Themes**	Two main "Tacit" and "Explicit" themes that complement and work together to advance or retreat macro-knowledge competitiveness
3	**Baskets**	Divide each theme into smaller units for top-down macro- knowledge progress monitoring
4	**Sub-Baskets**	Organise each basket to a more manageable units for closer diagnostics interpretability and control
5	**Indicators**	Brings each sub-basket to its major variables building blocks and source, for drill-down precise diagnostics, interpretability, diffusion and control
6	**Variables**	Brings each element to its key constituent elements for control and modifications

the World Bank - KAM tool (Chen and Dahlman, 2005; Chen and Kee, 2005) in order to provide a basic assessment of countries' environment, readiness, diffusion and absorption of knowledge on a macro scale. However, one of the major goals of the proposed framework is to unify the existing KBE and competitiveness related SCIs to create a bias free, unified with forecasting capability indicator to measure KBE competitiveness.

The suggested UKCI framework encompass a wide-range of objectives in its latitude and environment. The core UKCI index with its 2 themes, 8 major baskets and its sub-baskets is shown in Figure 4.6. Such divisions would allow for further mining of the framework to aid as an expressive and decisive decision making tool. For example, the major nodes of technology are renamed after further mining exposed a bigger hierarchy underneath it with the abbreviation of "ICT" which is an acronym for Information Communications and Technology, which is more accurate as it gives a wider coverage than the word "technology" and is widely used and known term. To further clear the justification, purpose and the advantages of such framework hierarchy. The subsequent section will present a brief descriptions of themes and its components.

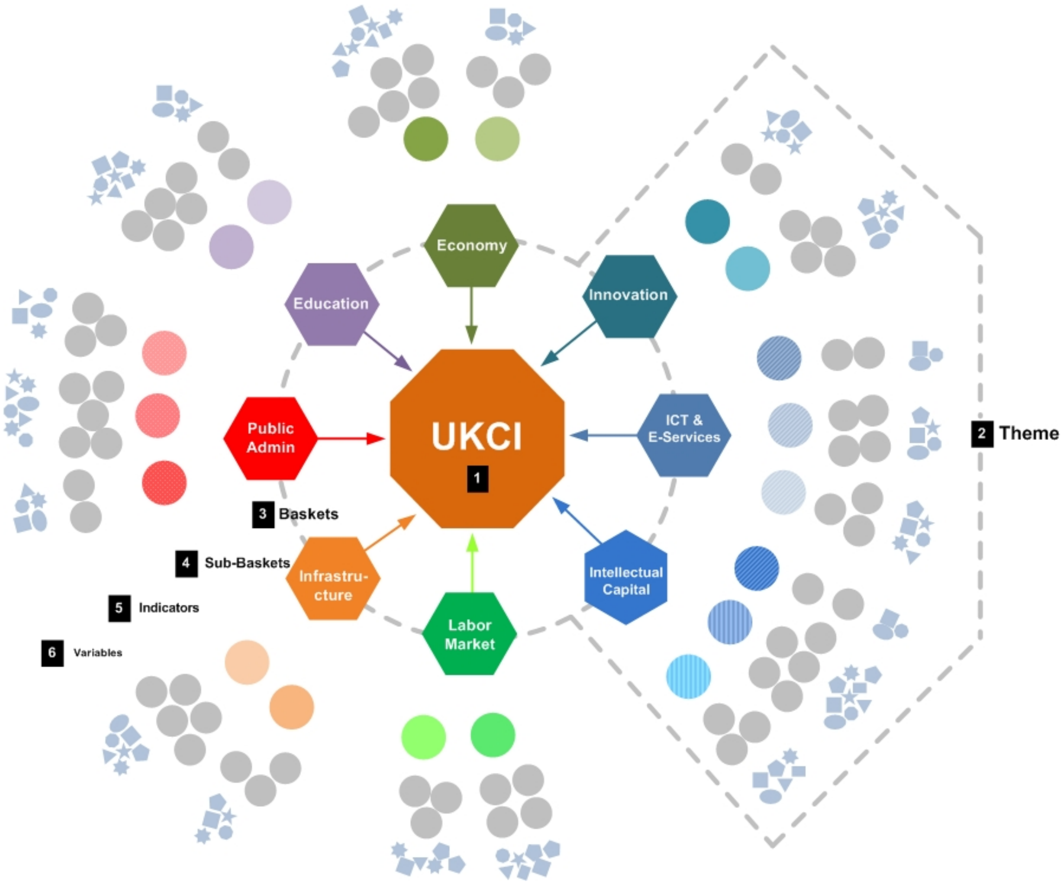


Figure 4.6: UKCI units of development and overall topology.

4.6.1 Tacit Theme

To the best of the author's knowledge there is no division or sub-composite index to measure the "tacit" aspects of the collective knowledge of a nation on a macro level. The existence of such indicator is perceived as beneficial as it would direct the attention of the decision maker and investors to which economy is tacit rich even though for example it is 'infrastructure' poor. The soft aspect, hence the 'tacit' theme of the knowledge economy is explained and defended in the previous chapter, when we devised the epistemology for a unified KBE indicator, as this study argues for the existence of such a theme which can be measured and monitored for the purpose of promoting KBE competitiveness, development and progress. The UKCI tacit theme consists of three major baskets as follows:

4.6.1.1 ICT and E-Services

This basket deals with ICTs which continue to penetrate countries in all regions of the world, as more and more people are getting connected. ICT plays an important role in improving a nation's capability to convert needs into applicable practices, which can aid in eradicating poverty and improving the delivery of education and healthcare (ITU, 2012). The ICT and E-Services basket consists of 5 sub-baskets that includes major ICT indicators such as ICT Environment, Access, Use, Price and the e-Services progress such as e-Government, e-Health etc.

4.6.1.2 Intellectual Capitals

This basket has emerged as a major addition compared to the existing WB- KEI composite index. The Intellectual Capital basket, comprises of 3 sub-baskets which contains the indicators of Human Capital, Structural Capital, and Relations Capital. The indicators for these three sub-baskets emerged mainly from the WEF-GCI and INS-GII indices. The benefits of keeping such measure has already been justified e.g. keeping an account of the Human Capital logical elements such as the stock of 'knowledge workers', the nation brain gain/drain and the business sophistication which could have the "spill-over" effect to benefit an economy. The Structural Capital on the other hand would measure research and development facilities, creative intangibles such as patents, inventions, copy rights etc. Relation Capital, however takes care of soft elements such as the networking, cooperation and collaborations between knowledge workers and institutions.

4.6.1.3 Innovation

Innovation is the most shared and agreed factor between the existing SCIs. Actually, the INSEAD Business School Global Innovation Index is an innovation-dedicated index. This basket contains 2 sub-baskets labelled as "innovation input" and "innovation output". The Input sub-basket measures the factors that would help create and enhance innovation, such as building scientific research centres, innovation linkages and the capacity of a society to quickly input and absorb knowledge, which is needed to advance innovation in a nation and to help create an innovative outputs. The output sub-basket on the other hand contains the results such as creative goods and services, knowledge impact and diffusion. The innovation output depends heavily on the innovation inputs of a nation, in other words, building advance research centres should help a country produce innovative technology or discover highly needed medication etc.

4.6.2 Explicit Theme

This theme deals with the classic or "hard" factors that contribute to progress and development on a macro level in and between nations. Keeping track and devising a tool to measure these elements is beneficial as it would direct the attention of the decision maker to the "explicit" aspects of the knowledge based economy. This intention is explained and defended in the previous chapter, when defining the epistemology for a unified KBE indicator, as this study argue for the existence of such theme which can be measured and monitored. This theme consists of five major baskets as follows:

4.6.2.1 Economy

The idea behind including the economy as a major pillar is explained by Stevens (1996) such that "investments in knowledge can increase the productive capacity of the other factors of production as well as transform them into new products and processes. Since these knowledge investments are characterised by increasing (rather than decreasing) returns, they are the key to long-term economic growth". This basket covers the micro and macro aspects of the economy through its two sub-baskets of Micro-Economy and Macro-Economy.

4.6.2.2 Education

The knowledge-based economy is characterised by the need for continuous learning of both codified information and the competencies to use this information. The relationship between learning and growth is well established by the different versions of the New Growth Theory which is accredited to Lucas (1988) and Romer (1991a). New growth emphasize the importance of formal learning process, namely education, research and learning-by-doing. The economic implications of learning-by-doing reflects the fact that an increase in the utilization of capital leads not only to a scale effect, but also to an increase in the knowledge used in production because of additional experience gained (Conceio et al., 1998).

4.6.2.3 Infrastructure

Having solid and reliable infrastructures such as roads, ports, electricity, hospitals, schools etc. creates a rich environment to build a healthy KBE. Such services give the people of a nation the opportunity to focus on improving their own abilities and knowledge rather than worrying about the daily hassles of life. Lack of, poor or deteriorated infrastructure creates a knowledge sink hole to the citizens of a society. This basket consists of 2 sub-baskets that covers the basic infrastructures such as roads, ports, healthy environment etc. and the advance infrastructure covers technological, scientific and energy etc. infrastructures in a country.

4.6.2.4 Labour Market

This basket relates to the labour forces and employments in a country and is not to be confused with the Human Capital which considers developing the individuals skills, while the labour market considers developing a good environment for HC to excel. Putting forward fair rules and regulations to protect the labour forces in a country helps increase creativity and productivity. Fighting high levels of unemployment spreads hope between the citizens and encourages them to piece together a bright future for all.

4.6.2.5 Public Administration

Government policies, law and order relating to technology, education and industry need new focus. For example, support to innovation will need to be on a state level to enhance the diffusion of new technologies and knowledge to all sectors. Government policies should promote lifelong learning and provide the adequate infrastructures such as schools, universities research centres, vocational training etc. to help create a highly skilled HC. Efficient and honest government is needed to protect the capitalised achievements and wealth of a nation, establish a safe harbour for creative and foreign investments, create and promote ethical practice

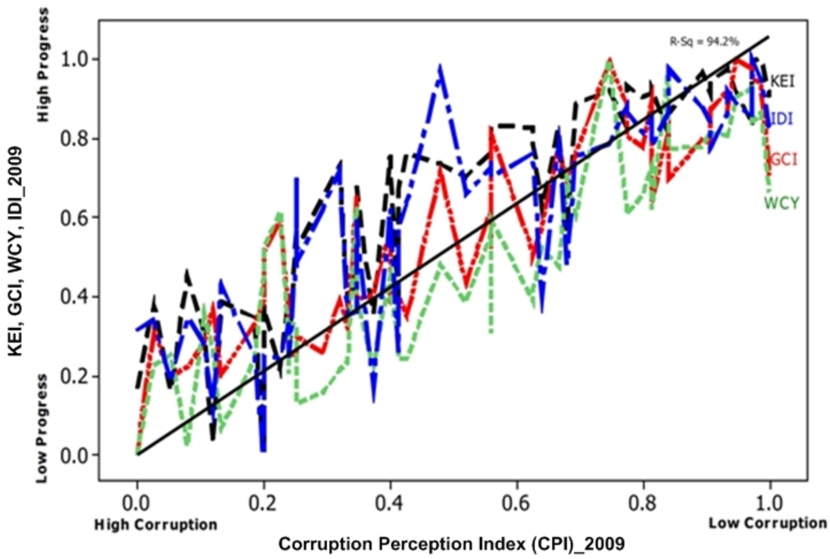


Figure 4.7: Corruption hinders knowledge progress, innovation and competitiveness.

and combating corruptions should also be a top priority of a Public Administration (PA). Good PA can provide the enabling conditions for achieving a healthy KBE environment through appropriate financial, competition, information and other related policies. Five sub-baskets have emerged from this large and important basket; government efficiency, political stability, law and order, social justice, ethics and corruption.

In light of this basket and its sub-baskets results, the first legitimate question that comes to mind is that what do ethics and corruption have to do with knowledge competitiveness and progress in a society? This is a valid question and by answering this, it may contribute to the reliability and justification of the achieved overall taxonomy. To answer the question, a simple correlation and regression between the final scores given to 57 countries in 2009 by four of the earlier mentioned composite indicators (KEI, GCI, WCY and IDI) was put against the score for the same countries of the main global Corruption Perception Index, by Transparency International (Lambsdorff, 2003). This simple yet effective test would show that there is a relation between the levels of corruption and how it affects macro-knowledge progress, competitiveness, innovation, or even technology advancement in a nation. Figure 4.7 presents the scatter plot for the Pearson Cor-

Table 4.3: Extracted elements for macro-knowledge competitiveness.

	Themes	
	Explicit	**Tacit**
	Economy	**ICT and e-Services**
	Macro-Economy	ICT Environment
	Micro-Economy	ICT Readiness
	Education	ICT Use
	Basic Education	ICT Price
Baskets	Higher Education	E-Services
&	**Infrastructure**	
Sub Baskets	Basic Infrastructure	**Intellectual Capital**
	Advance Infrastructure	Human Capital
	Public Administrations	Structural Capital
	Government Efficiency	Relations Capital
	Political Stability	**Innovation**
	Law and Order	Innovation Input
	Social Justice	Innovation Output
	Ethics and Corruption	
	Labour Market	

relation and the regression line of the main Corruption Perception Index against the KBE and competitiveness indices KEI, GCI, WCY and IDI. This shows a strong positive relation between the level of corruption in a country and the level of progress in knowledge economy ($R^2 = 94.2\%$). This result asserts the fact that a high level of corruption would definitely hinder a country progress in many aspects. This test also strengthens the credibility of the produced taxonomy.

4.6.3 UKCI Concept Map

The UKCI baskets, sub-baskets, indicators and the composite indicator sources can be traced through a detailed, semi-detailed table and fully detailed visual concept map of the assembled framework. Table 4.3 itemises the final devised themes, emerged baskets and sub-baskets as a direct result of the FPKM process of the targeted knowledge and competitiveness indicators. Figure 4.8 presents the UKCI detailed visual concept map to bring each basket to its original sources. This concept map would document the building and brings each sub-basket to its constituent elements and sources to serve in future enhancements.

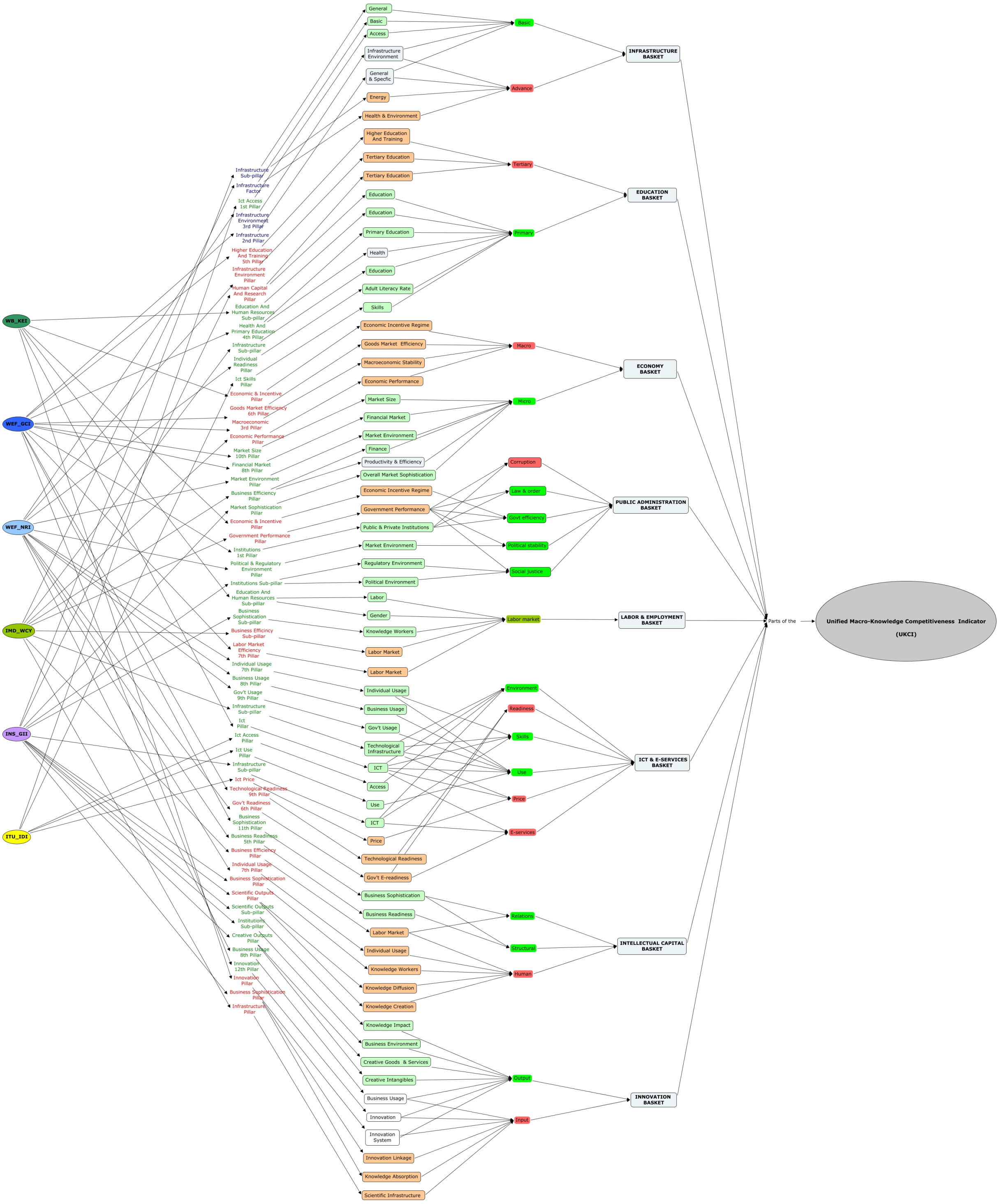


Figure 4.8: UKCI concept mapping.

4.7 Summary

In this chapter an intelligent method is proposed and presented to create a qualitative taxonomy - without any "experts" interventions - to precisely acquire and remap several hundred macro-knowledge and competitiveness related variables for the purpose of creating a unified KBE competitiveness framework. This was successfully achieved through the use of a hybrid model of exact and fuzzy text matching. One of the advantages of SCIs is that it takes into considerations the interconnections between the defined themes, baskets, sub-baskets and indicators. Thus, the 2 themes, 8 major baskets, 22 sub-baskets and its 80 indicators relevant to KBE competitiveness and progress are spotted, classified, remapped and renamed thorough an interpretation of the main issues in the original KBE, competitiveness, ICT, Innovation and progress composite indicators. These are then combined to create the UKCI qualitative taxonomy and concept map.

Having established the qualitative taxonomy for the proposed UKCI, in the following chapter presents the rationalisation of the actual data that have been collected to represent their particular themes, baskets, sub-baskets and indicators. Moreover, the methods and techniques utilised by this research is presented, as well as a comparative review of the computational intelligence vs. the statistical methods used in the field of creating composite indicators.

Chapter 5

Data and Methods of Development

5.1 Introduction

In the previous chapter, the qualitative taxonomy underpinning the structure and concept for the *UKCI* was introduced. The emphasis for this chapter is to present the details of the methods used to quantitatively and empirically construct and assess the UKCI indicator. The methods explained in this chapter answers to the stated objectives of the current study, which is to evaluate and propose an alternative methods to the measurement of KBE competitiveness which integrate the strengths and resolve the shortcomings of the reviewed approaches, to establish a unified, non-bias, intelligent and robust Synthetic Composite Indicator.

This chapter will begin with a brief review of the datasets which was collected and arranged based on the earlier composed qualitative taxonomy. These datasets are utilised for analysis, particularly the indicators type, variables contingency, country selections, and time periods availability. This will be followed by the procedures and techniques usually used for developing composite indicators. This includes, but not limited to, data treatments and analysis such as multivariate analysis, correlations, normalisation, missing data imputations, outliers detection techniques, weighting, aggregation and robustness analysis. This chapter also propose alternative CI techniques specifically to impute missing data, weight and aggregate variables. The objective here is to guarantee a coherent reliability occurrence between the qualitative taxonomy and the quantitative *UKCI* indicator. Additionally, this chapter will provide a hybrid techniques in the area of advance econometrics such as Panel data analysis and ANN, leading to a list of potential forecasting techniques that will be used later in Chapter 7 to predict and forecast the future directions of a certain KBE regardless of missing or limited data availability.

This chapter is organized as follows; in Section 5.2 a brief review of the numerical data that was collected and arranged based on the composed qualitative taxonomy and the year periods is presented. Section 5.3 gives an overview of the different methods to be employed. Section 5.4 introduces the data treatments and handling. In Section 5.5 and Section 5.6 respectively , the different statistical weighting and aggregation methods are presented. The employed CI techniques are presented in Section 5.7. The statistical and CI techniques used to impute missing data are presented in Section 5.8. Section 5.9 presents the methods of predictions and forecasting. The robustness and validation methods used to test the proposed models are explained in Section 5.10. The chapter summary is presented in Section 5.11.

5.2 Data Collection

Many simple and composite indicators are developed to measure all aspects of progress and developments on a micro and macro level. However, this study is guided by the focus on indicators, according to their relevance, availability and wide usage. In addition, the datasets used in this study are collected based on the qualitative taxonomy created in the previous chapter. These datasets are freely and readily available from the annual reports of the organizations mentioned earlier. Each set contains various numbers of economies as reported by the issued entities. Furthermore, this study intends to using real variables to form its datasets to accurately test the validity of the proposed methods. However, the idea of measuring KBE competitiveness and progress is a new realised concept, therefore serious and full data for multiple years is problematic especially for a large number of the developing and underdeveloped economies. Also, the time periods covered by these composite indicators vary, and even the range and economies (countries) included in these indices could be different. Nevertheless, this research established and kept a neat collection of the different used datasets which can be downloaded with permission from the author [1]. Some of these data sets dates back to 2007, however, It was not possible to acquire and fully match more than 57 countries for three consecutive years 2009, 2010, and 2011. Table 5.1 lists a summary of the main sources of KBE indicators.

To measure and forecast KBE progress in a certain economy, the data can be treated as cross sectional, where many countries' progress are observed over the same point in time. Otherwise, it can be treated as time-series data, where a specific country is followed over the course of time. However, after studying the characteristics of the collected data set it has been noticed that the level of change in a certain knowledge economy does not happen overnight or even from one year to another, rather it is a slow progress through the accumulation over the years. To confirm this fact, a score correlation has been conducted for each of the four selected indicators across the years. The individual score correlation results are presented in Table 5.2, which shows a correlation that ranges between 95% and 99%, asserting an important fact that knowledge economy progress and capability signifies a deep infrastructural development, therefore substantial change in the ranking between economies does not happen in a short term period. Thus, one can assume that the change in the level of KBE for a certain economy for the next year will depend upon the change in the reported scores of the above mentioned indicators and on the level of the reported progress from the previous years.

[1] www.alshami.info

Table 5.1: Selected knowledge economy indicators.

Indicator	Organization	Years	Eco. Counts	Score Scale
KEI	WB	1995-2012	146	0-10
GCI	WEF	1979-2012	134	1-7
NRI	WEF	1979-2012	134	1-7
WCY	IMD	1989-2012	59	0-100
GII	INS	2009-2012	125	0-100
IDI	ITU	2002-2012	159	0-10

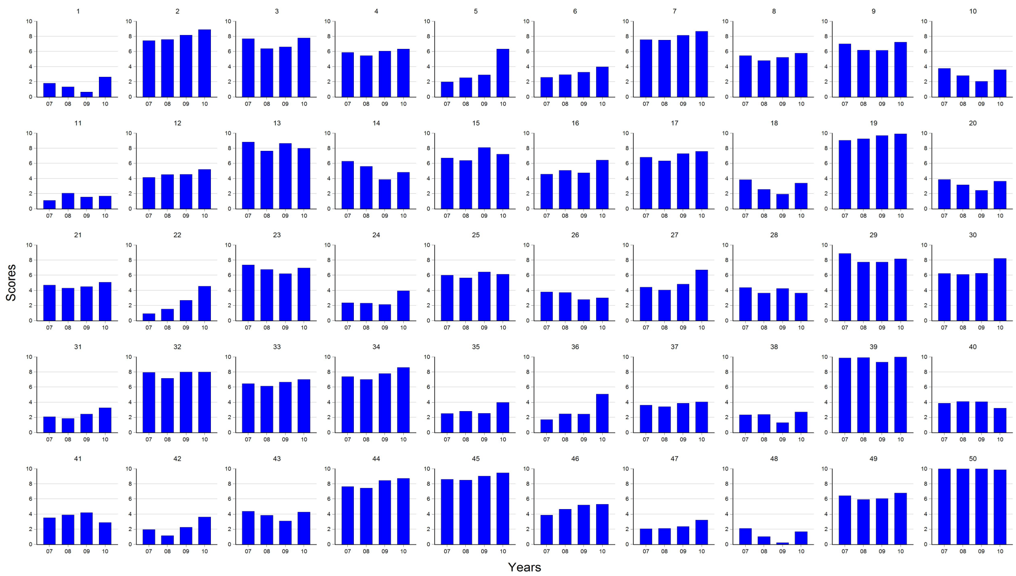


Figure 5.1: WCY scores from 2007-2010 for 50 random economies.
Data Source:(IMD, 2011)

Tracking the changes within a certain knowledge economy and between other observed KBEs can be captured by following the scores issued by the above mentioned organisations. This can be charted and then visualised if presented together in separate bar charts. Figure 5.1 illustrate the WCY scores to show how one can visualise many different KBEs at once as they progress over the years, where each chart represent a certain economy scores tracked over four consecutive years.

Table 5.2: Individual indicator score correlation for three consecutive years 2009, 2010 and 2011.

KEI	**2009**	**2010**	**2011**	**IDI**	**2009**	**2010**	**2011**
2009	1.0			**2009**	1.0		
2010	0.99	1.0		**2010**	0.99	1.0	
2011	0.98	0.99	1.0	**2011**	0.99	0.99	1.0
GCI	**2009**	**2010**	**2011**	**WCY**	**2009**	**2010**	**2011**
2009	1.0			**2009**	1.0		
2010	0.99	1.0		**2010**	0.98	1.0	
2011	0.98	0.99	1.0	**2011**	0.95	0.97	1.0

On the other hand if evaluating a specific or few KBE is the main concern, the radar chart could be useful for visualisation and thus evaluation and/or prediction of competitiveness of countries in time (dynamic process). In Figure 5.2, where data for Economy 1 and 2 are visualised for the years 2007, 2008, and 2009 presented in different style dashed lines, as an example for further possible evaluation and decision making.

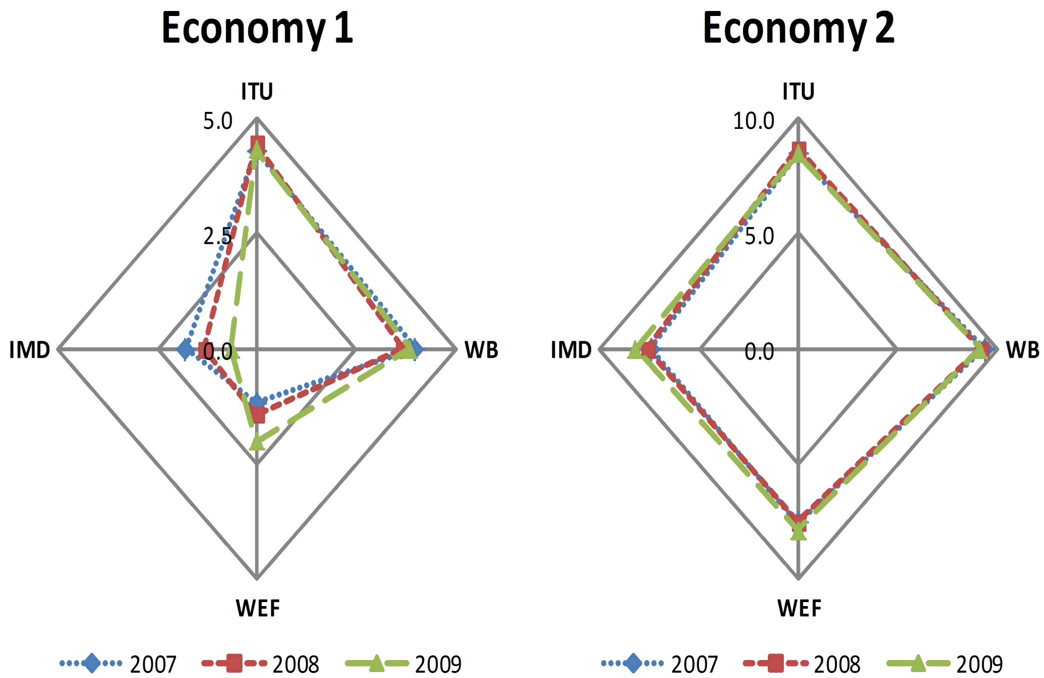


Figure 5.2: Two random economies on a radar visualisation from four selected indicators.

5.3 SCIs Methods of Construction

In this section, the SCI methods of construction used in this thesis are reviewed. These techniques will help to provide information related to the types and which methods would be appropriate to unify, predict and forecast the progress and competitive behaviour of KBE to build the suggested *UKCI*. More details about the results of the unification and forecasting process which involved statistical and computational intelligence for weighting, aggregation and forecasting techniques are presented in later chapters. In this chapter however, the techniques to collect, treat, weight, aggregate and forecast the dataset are reviewed. In brief the investigated techniques are: multivariate analysis, such as PCA and Cronbach's Alpha (C-Alpha); standardization methods, such as Z-Score and Min-Max; correlation and association methods such as Pearson Correlation Coefficient and Chi Square. For missing data analysis, two special Fuzzy C-Means techniques that is, the Optimal Completion Strategy (OCS) and the Nearest Prototype Strategy (NPS), from the CI side to impute missing values. The results are compared against four statistical imputation techniques namely; the Expectation Maximisation, Multiple Imputation, Nearest Neighbour and Regression. For clustering the Euclidean distance measure and Time Distance techniques were explored; For the weighting and aggregation methods (unifications), two CI methods were investigated; FCM and Vector Quantisation (VQ) were compared against major statistical methods such as PCA/FA to weight and the Geometric Mean to aggregate. Finally, for prediction and forecasting, a hybrid approach is used. It consists of Panel Data: Time Series Cross Sectional (TSCS) and ANN were compared against pure ANN, TSCS and Linear Multiple Regression methods.

5.4 Data Treatments Methods

Over the last few decades, national and international agencies produce a large amounts of statistical data and composite indicators to measure various progress and competitiveness domains. Even though these raw data allow for high level evaluations, usually the data reported for comparison between countries is not systematically harmonised, which makes evaluations challenging and can hinder the usefulness of the available data. Hence, the underlying nature of the data used for any study has to go through rigorous procedures before the development of SCI. Such procedures are known as data treatment techniques where the data from different sources, angles and perspectives has to be carefully collected, studied, transformed, scaled and treated of an anomalies, outliers, or missing data and then summarised for easy calculations, evaluations and/or visualisations. Skipping through these important steps can lead to the production of "information poor" and "positively biassed" indicators which can lead to indices that confuse, mislead and overwhelm public officials, and citizens. In the following sections the methods to be used in this thesis for the purpose of constructing the proposed *UKCI* is explained. In general the notation that is adopted throughout this chapter is as follows unless it is stated otherwise.

X^t_{ic} : Represent the value of variable i for country c at time t, with $i = 1, \ldots, M$ and $c = 1, \ldots, N$

X'^t_{ic} : Is the normalised value of indicator i

w_{ri} : weight associated to sub-indicator i, with $r = 1, \ldots, R$

SCI^t_c: value of the composite indicator for country c at time t.

5.4.1 Multivariate Analysis

The goal of multivariate analysis is to investigate the inherent structure in the indicators set to reveal how different variables change in relation to each other and how they are associated.

The first step in building a SCI is to decide whether the structure of the SCI is thoroughly described and if the variables are adequate or suitable to measure the phenomenon under investigation. This can be decided based on experts' opinion or based on the arithmetical formation of the dataset, to provide a sound and defensible dataset. For example PCA or measurement of internal consistency (reliability) such as Cronbach's Alpha can be used to investigate whether the different used variables are statistically well balanced to make the composite desired indicator. If this is not true, amendment of the used variables might be considered (OECD, 2008a).

Principal Components Analysis

Principal Components Analysis (PCA) is one of the multivariate and inputs reduction method. The goal of PCA is to reveal how different variables are associated and how they change in relation to each other (endogenous vs exogenous variables). PCA is useful when we have two or more variables, and believe that there is some redundancy in those variables. In this case, redundancy means that some of the variables are correlated with one another, possibly because they are measuring the same construct. Because of this redundancy, it should be reasonable to reduce the observed variables into a smaller number of principal components "artificial variables" without much loss of information. The dimensionality reduction mechanism of PCA is to explain the variance of the observed data through a few linear combinations from the original data (Jolliffe, 2002).

The objective of PCA is to take m number of variables $x_1, x_2, ..., x_m$, and spot the linear combinations of these to produce uncorrelated principal components $P_1, P_2, ..., P_m$ as follows:

$P_1 = w_{11}x_1 + w_{12}x_2 + ... + w_{1m}x_m$

$P_2 = w_{21}x_1 + w_{22}x_2 + ... + w_{2m}x_m$

...

$P_i = w_{i1}x_1 + w_{i2}x_2 + ... + w_{ij}x_j + ... + w_{im}x_m$

...

$P_m = w_{m1}x_1 + w_{m2}x_2 + ... + w_{mm}x_m$

the weights w_{ij} -(also called component or factor loadings) applied to the variables x_i are chosen so that the principal components P_i satisfy the following conditions:

- they are uncorrelated (orthogonal);
- the maximum possible proportion of the variance of the set of xs, will be accounted by the first principal component, the maximum of the remaining variance will be accounted for by the second principal component, and so on until the all the remaining variance not accounted for by the preceding components will be absorbed by the last principal components.
- the factor loadings w_{ij} related to the variable x_j should sum up to 1.

In brief, PCA involves finding the Eigenvalues $\lambda_j, j = 1, 2, ..., m$, of a sample covariance matrix VM

$$V = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1m} \\ v_{21} & v_{22} & \dots & v_{2m} \\ \dots & & & \\ v_{m1} & v_{m2} & \dots & v_{mm} \end{bmatrix} \tag{5.1}$$

where the variance of x_i and v_{ij} is the covariance of variables x_i and x_j and represented by the diagonal line v_{ii}. The sum of the diagonal values equals the eigenvalues of V. There are m eigenvalues, which are the variances of the principal components. This means that, the total of the principal components variances is equivalent to the total of the original variables variances and as follows:

$$\lambda_1 + \lambda_2 + ... + \lambda_m = v_{11} + v_{22} + ... + v_{mm}. \tag{5.2}$$

To prevent a certain variable from overriding the position over other variables on the principal components, it is suggested to first normalise the variables using the standardisation or (z-score) normalisation technique. As a result, all variables in the dataset will have equal means of "zero" and variances of "one" (Sarle, 1994). PCA was employed in this study to serve three purposes: first, to test if the variables could be reduced. Second, to reduce the number of indicators to a smaller subset. Third, to foresee the possibility of filtering out the trivial components, before we use it. The trivial components usually act as noise and could stand on the way of getting a sound and meaningful clustering results.

5.4.2 Data Normalisations

Normalisation usually is used to transform different measurement units into the same unit, so they can form a clear comparable elements, and to avoid problems in mixing measurement units (e.g. money, talent, skills) (Freudenberg, 2003). For such cases it is recommended to use the standardisation or (z-score) normalisation technique. As a result, all variables in the dataset will have equal means of "zero" and a standard deviations of "one" (Sarle, 1994). However, given that the selected indicators use different score scales and units in the collected dataset, the data requires transforming or adjustments to become comparable and to convert the different ranges of indicators into a unified range. Hence, for this case the issue is not only the use of different measurement units, but also the difference in scores scale ranges. So to unify the score ranges between the different selected indicators, it is therefore suggested to use the Min-Max normalisation techniques, which can be applied by taking all the different scores ranges collected in the data set and transforming it to a value between 0 and 1, where the lowest (min) value is set to 0 and the highest (max) value is set to 1. These normalisation methods can be expressed as follows:

$$X'_i = \frac{x_i - minA}{maxA - minA} \times (new\_maxA - new\_minA) + new\_minA \tag{5.3}$$

where x'_i is the normalised score, x_i is the actual score, $minA$, $maxA$ are the minimum and maximum values of the scores range within index A.

In the cases where a high value implies inferior result (e.g., ICT Price, corruption, tariff rate, unemployment), this study resort to normalization formula that, in addition to converting the series into a $[0 - 1]$ range, inverts it, so that 0 implies poor and all the way to 1 as the top possible performances:

$$X'_i = \frac{x_i - minA}{maxA - minA} \times (new\_minA - new\_maxA) + new\_maxA \tag{5.4}$$

5.4.3 Measures of Correlation and Association

To investigate the relation between numerical variables, the data observed can be tested using correlation and contingency analysis. These tests allow us to test if the relation between variables is strong enough to indicate whether the produced results are significant. Two measures are summarised in the following sub-sections.

5.4.3.1 Pearson Correlation Coefficient

Pearson Correlation Coefficient ,r, can be used to test relations between different variables. It can be calculated by dividing the covariance of two variables x_1 and x_2 by the product of the standard deviations for both variables. r can be expressed as follows.

$$r_{x_1 x_2} = \frac{Cov_{x_1 x_2}}{(\sigma x_1 \times \sigma x_2)} \tag{5.5}$$

$$Cov_{x_1 x_2} = \frac{\sum (x_{1i} - \bar{x_1})(x_{2i} - \bar{x_2})}{(n-1)} \tag{5.6}$$

where σ is the standard deviation of the variables x_1 and x_2, $\bar{x_1}$ and $\bar{x_2}$ are the mean of the sample variables of x_1 and x_2 values.

5.4.3.2 Chi-Square Based Measures

One way to determine whether there is a statistical relationship between two variables is to use the chi-square χ^2 test for independence. A cross classification table is used to obtain the expected number of cases under the assumption of no relationship between the two variables. Then the value of the chi-square statistic provides a test whether or not there is a statistical relationship between the variables in the cross classification matrix (Mantel, 1963). The following formula sums the procedure as follows:

$$\chi^2 = \sum \left[(O - E)^2 / E \right] \tag{5.7}$$

where O is the observed frequencies and E is the expected frequency. The expected frequency can be calculated using the following equation:

$$E = \frac{\sum Row \times \sum Column}{\sum Overall} \tag{5.8}$$

5.4.4 Cluster Analysis

Cluster Analysis (CA) is the process of finding similarities between homogeneous characteristics found in a data set. Hence, distinct or alike data points could be mapped together, based on the distance between the data points, where large distance means weak cluster, and small distance means strong similarities. The goal of CA is to decrease the dimensionality of a dataset by surfacing the unseen similarities and dissimilarities. Cluster Analysis is useful in that regard, and will be utilised in different sections of this study. Distance measures includes Euclidean (geometric) vector space and non-Euclidean, however, the most common is the Euclidean because they perform well in multi-dimensional space. The distance between data points reflect the detected similarities or dissimilarity, for example the distance between two points (X_1, X_2) over N_d dimensions can be calculated using Euclidean Distance (ED) formula:

$$D(X_1, X_2) = \sqrt{\frac{\sum_{i=1}^{N_d}(X_{1i} - X_{2i})^2}{N_d}} \tag{5.9}$$

5.5 Statistical Weighting Methods

A weight giving to a certain variable can highly influence the outcome of SCI and the overall country rankings. There are many traditional statistical weighting techniques for example, Equal Weighting (EW), Data Envelopment Analysis (DEA), Factorial Analysis (FA), and Benefits of the Doubt (BOD). There are also experts and, or stakeholders participatory methods such as Analytic Hierarchy Processes (AHP), Conjoint Analysis (COA) and Budget Allocation Processes (BAP). Similarly, the Arithmetic Mean (AM), Geometric Mean (GM), Additive Rules (AR) etc. are regularly used for the purpose of aggregating the variables to form a single value, hence a "composite index" (OECD, 2008a). This research study steered away from using any of the participatory methods, which are solely dependant on the opinions and judgement of the surveyed people or "experts" for developing micro or macro measurements. Such methods suffer from the subjectivities, possible biases and personal intuition of opinions for weight settings. Furthermore, such methods requires allocations of resources to hire credible and trustworthy "experts", which is beyond the scope, intentions and limits of this research study. For the purpose of comparisons and proof we used two of the most widely used statistical weighting methods which are explained below.

5.5.1 Equal Weighting

When all variables are giving equal weighting it may happen that by combining variables with high degree of correlation, one may introduce an element of double counting into the index. If two collinear indicators are included in the composite index with a weight of w_1 and w_2, then the unique dimension that the two indicators measure would have weight $(w_1 + w_2)$ in the composite indicator. Therefore, it is recommended to test the correlation between the indicators using Pearson Correlation Coefficient, and choosing only indicators exhibiting a low degree of correlation or adjusting weights correspondingly, e.g. giving less weight to correlated indicators. Furthermore, minimizing the number of variables in the index may be desirable on other grounds such as transparency and interpetability. It should be mentioned that there will always be some positive correlation between different measures of the same aggregate. Thus, a rule of thumb should be introduced to define a threshold beyond which the correlation is a symptom of double counting. On the other hand relating correlation analysis to weighting could be dangerous when motivated by apparent redundancy (OECD, 2008a; Grupp and Schubert, 2010).

5.5.2 PCA and Factor Analysis Weighting

Principal Component Analysis, and more specifically Factor Analysis groups together individual indicators which are collinear to form a composite indicator that captures as much as possible of the information common to individual indicators. Each factor (usually estimated using PCA) reveals the set of indicators with which it has the strongest association. The idea behind PCA and FA is to account for the highest possible variation in the indicator set using the smallest possible number of factors (Johnson and Wichern, 2007).

5.6 Statistical Aggregation Methods

There are couple of aggregation techniques, when developing composite indicators. However, one of two major aggregation techniques is often used:

- Linear Aggregation (LA): It is a summative aggregation method usually used with equal weighting criteria. LA can be formulated using the following formula:

$$SCI_c = \sum_{q=1}^{Q} w_i x_i' c \tag{5.10}$$

- Geometric Mean (GM): It is a multiplicative aggregation technique and can be performed using the GM equation, which is stable and highly recommended for aggregating SCI, and such technique is widely studied in fuzzy set theory (Zimmermann and Zysno, 1983). The GM can be expressed using the following equation:

$$SCI_c = \prod_{i=1}^{I} x_{i,c}^{w_i} \tag{5.11}$$

where SCI is the aggregated composite score for a certain country c, x is the value of each variable or indicator i for country c raised to the power w, which is the weight assigned for each variable or indicator i.

5.7 Computational Intelligence Techniques

Computational Intelligence (CI) methods are becoming popular for their precision modelling, clustering, predictions and trend analysis. Some methods could be used as an alternative to the statistical techniques. The proceeding sections presents a review of the methods used in this study.

5.7.1 Fuzzy c-Means Algorithm

Fuzzy c-Means (FCM) is a clustering algorithm introduced by Dunn (1973) initially and improved by Bezdek et al. (1984). FCM stems from the famous K-means algorithm, but it differs in that the data point has partial membership in a cluster, with grades between 0 and 1. Therefore, FCM allows one piece of data to be a member of two or more clusters according to its degree of membership, which is determined based on the distance (usually the Euclidean) between a data point and the cluster centre. At each iteration, an objective function is minimized to find the best location for the clusters and its values are returned in objective function.

Fuzzy clusters can be characterised by class membership function matrix, and cluster centres are determined first at the learning stage, and then the classification is made by the comparison of euclidean distance between the incoming features and each cluster centre.

For a data set represented as $X = \{x_1, x_2, \ldots, x_j \ldots, x_n\} \subset R^s$ into c clusters, where $1 < c < n$; the fuzzy clusters can be characterized by a $c \times n$ membership function matrix U, whose entries satisfy the following conditions:

$$\sum_{i=1}^{c} u_{i,j} = 1, \qquad j = 1, 2, \ldots, n \tag{5.12}$$

$$0 < \sum_{j=1}^{n} u_{i,j} < n, \qquad i = 1, 2, \ldots, c \tag{5.13}$$

where $u_{i,j}$ is the grade of membership for x_j data entry in the ith cluster. Cluster centres are determined initially at the learning stage. Then, the classification is made by comparison of distance between the data points and cluster centres. Clusters are obtained by the minimisation of the following cost function via an iterative scheme.

$$J(U, V) = \sum_{j=1}^{n} \sum_{i=1}^{c} (u_{i,j})^2 \left\| x_j - v_i \right\| \tag{5.14}$$

where $V = \{v_1, v_2, \ldots, v_i, \ldots v_c\}$ are c vectors of cluster centres with v_i representing the centre for ith cluster.

To calculate the centre of each cluster, the following iterative algorithm is used.

1. Estimate the class membership U.

2. Calculate vectors of cluster centres

$V = \{v_1, v_2, \ldots, v_i, \ldots v_c\}$ using the following expression:

$$v_i = \frac{\sum_{j=1}^{n} (u_{i,j})^2 x_j}{\sum_{j=1}^{n} (u_{i,j})^2} \qquad i = 1, 2, \ldots, c \tag{5.15}$$

3. Update the class membership matrix U with:

$$u_{i,j} = \frac{1}{\sum_{r=1}^{c} \left(\frac{\|x_j - v_i\|}{\|x_j - v_r\|} \right)^2} \; i = 1, \ldots, c; \;\; j = 1, \ldots, n \tag{5.16}$$

4. If control error (defined as the difference between two consecutive iterations of the membership matrix U) is less that a pre-specific value, then the process can stop. Otherwise process will repeat again from step 2.

After a number of iterations, cluster centres will satisfy the minimisation of the cost function J to a local minimum.

5.7.2 Vector Quantization

Vector Quantization (VQ) is a classic technique from signal processing and usually used for data compression, to recode data into more reduced forms. One such technique is which maps groups of input symbols, called vectors, onto a small set of vectors, called the “codebook”. Each vector in the codebook is a codeword (Zheng et al., 1997). VQ is an approximator, similar to that of rounding-off to the nearest integer (Linde et al., 1980).

To illustrate the concept let us assume that there is a training sequence consisting of n source vectors $T = x_1, x_2, ..., x_n$. The training sequence can be obtained from some large database. n is assumed to be sufficiently large so that all the statistical properties of the source are captured by the training sequence. It is assumed that the source vectors are k-dimensional, e.g.,

$X_n = \{x_n 1, x_n 2, \ldots, x_n k\}, n = 1, 2, ..., N$

let n be the number of code-vectors and let $C = \{C_1, C_2, ..., C_n\}$, for the codebook. Each code-vector is k-dimensional, e.g.,

$c_n = (c_n 1, c_n 2, ..., c_n k), n = 1, 2, ..., N.$

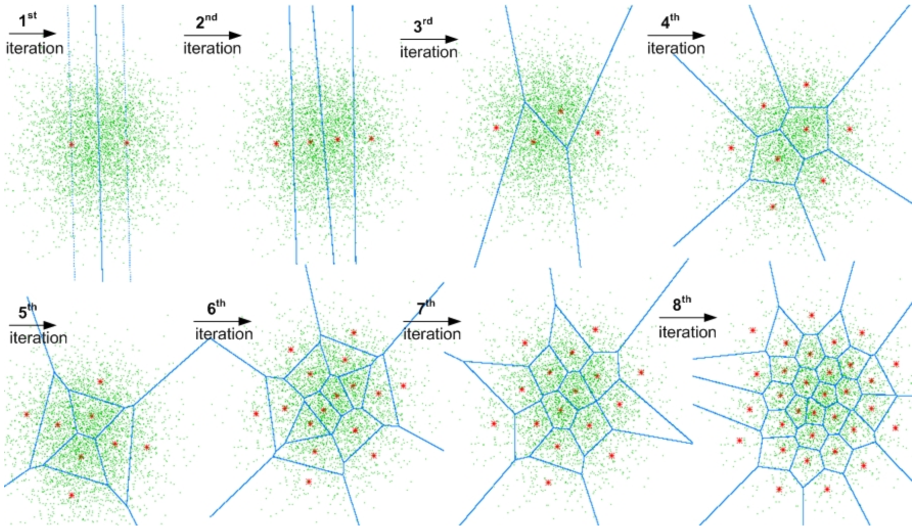


Figure 5.3: Illustration of the LBG LVQ clustering process.

Let S_m be the encoding region associated with code-vector c_m and let $P = \{S_1, S_2, ..., S_m\}$ denote the partition of the space. If the source vector x_m is in the encoding region S_m, then its approximation represented by $Q(x_n)$ is C_m:

$Q(x_n) = c_m$, if $x_n \in S_m$.

(Gray, 1984); (Wu and Guan, 1994).

LBG VQ Design Algorithm

The main purpose of utilising the Linde Buzo and Gray Vector Quantization (LBG VQ) algorithm is because of its ability to learn to detect and produce the centroid, the codebooks and codevectors between given scores. The LBG VQ algorithm is an iterative method which alternatively solves the above two optimality criteria (Linde et al., 1980). The algorithm requires an initial codebook. This initial codebook is obtained by the splitting method. In this method, an initial codevector is set as the average of the entire training sequence. This codevector is then split into two as the initial codebook. The final two codevectors are splitted into four and the process is repeated until the desired number of codevectors is obtained (Ramamurthi and Gersho, 1986). Figure 5.3 illustrates the iteration and splitting process for some randomly generated data points. The LBG VQ algorithm is summarized in Appendix A.

5.8 Missing Data Imputation Methods

Mining national and international statistical data for the purpose of developing SCI is usually associated with missing data. This major problem jeopardises the reliability of any index, as missing data can produce biased estimates, deform outcomes and void rankings.

5.8.1 Fuzzy c-Means Strategies

The limitation of the FCM algorithm is that it cannot be directly applied to incomplete datasets, because it needs to reference each vector for each value in a dataset. However, Hathaway and Bezdek (2001) proposed four techniques which can be integrated with FCM to allow it to accept and cluster incomplete datasets. These methods can be summarized as;

- Whole Data Strategy (WDS): In this method, all records with missing values are removed and the original FCM is applied to the full remaining dataset. If the dataset contains a high percentage of missing values, then it is not preferred, because dropping the missing values could result in loss of valuable and critical information,

- Partial Distance Strategy (PDS): In this approach, the partial distance is calculated using available features,

- Optimal Completion Strategy (OCS): In this approach, the missing values are viewed as extra variables to be optimised and therefore it imputes missing values at each iteration cycle till it reaches the best estimates,

- Nearest prototype Strategy (NPS): Is a slight modification of OCS in that it calculates the partial distances, and missing values are replaced by the their nearest prototype counterparts during each iteration.

According to the authors in (Hathaway and Bezdek, 2001), WDS and PDS are faster to terminate, but when it comes to accuracy and misclassification errors, the OCS and NPS methods were proved by theory and experiments as superior over the first two methods. Therefore, this study will compare the fuzzy clustering in

OCS and NPS strategies against four statistical imputation methods in an effort to accurately substitute missing scores when producing the *iSCI*.

5.8.2 Statistical Strategies

Different imputation methods are used to help in the substitution of missing data especially for the under or developing nations where key development data are neither collected or reported. There are a number of approaches to treat missing data, however, over time several techniques and methods have slowly progressed and some older methods have been abandoned. Nevertheless, many international organisations still resort to traditional statistical methods when imputing the missing data, such as case wise deletion where the missing values are simply dropped from the dataset.

- Mean Substitution (MS), is used when the sample mean of the present values for a certain indicator is calculated to substitute the missing values. MS is not recommended as it reduces the variance, which enlarges the error in any further analysis.

- Regression (REG) is one of the most widely used methods. In the regression method, missing values are imputed by predicted values based on the least square method, by forming a multiple regression equation where the indicator with the missing values is the dependent variable of the model. The other individual indicators form the independent variables of the model. In general the aim is to fit a model of the general regression form as follows:

$$\begin{aligned} Y_{ij} &= \alpha_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_j x_{ji} + ... + \beta_n x_{ni} + \varepsilon_i \qquad (5.17) \\ & n = 1, 2, \ldots, N; i = 1, 2, \ldots, M \end{aligned}$$

  where Y_{ij} is the dependant variable (to be imputed), x_{ji} are the independent variables, β is the coefficient vector and x_i is the i^{th} observation on j explanatory variables. The subscript i denotes entities, and it represent the countries: $n = 1, \ldots, N$; and ϵ_i is the random error term.

- Expectation Maximisation (EM), which was contributed by (Dempester et al., 1977). It consists of two steps: the expectation and the maximisation step, within an iterative process starting with some initial guess and an algorithm epoch, each step is completed once. At the expectation step the distribution of the missing points is determined from the known points of the presented variables using some other methods such as multiple regression. At the maximisation step, parameters with "maximum likelihood" are recalculated, which requires taking the derivatives of the likelihood function with respect to all the unknown parameters, assuming a correct distribution obtained in the expectation step. The maximum likelihood works with the relationship between unknown parameters of the data model and the missing data, where such unknown parameters would aid in obtaining robust prediction for the missing values (Neal and Hinton, 1998).

- Nearest Neighbour (NN), is usually called (hot-deck) imputation technique. This method fills in missing scores for a given country with available scores drawn from 'similar' or nearest neighbour countries, by calculating the distance (e.g. Euclidean or Manhattan) (Little and Rubin, 2002).

- Multiple Imputation (MI), is an iterative process which employ methods such as regression or Markov Chain Monte Carlo (MCMC), to impute multiple complete datasets using randomised techniques that reflects uncertainty, then it uses the multiple imputed datasets for the analysis. One of the most general models usually used is the MCMC method, which is a sequence of random variables in which the distribution of the actual element depends on the value of the previous one (Schafer, 1997).

5.9 Predictions and Forecasting Methods

As mentioned in the literature review, there are many forecasting techniques which are categorized into quantitative, time series, econometrics, judgemental, naïve and computational intelligence techniques. The subsequent sections will explain the methods of prediction and forecasting used in this thesis.

5.9.1 Multiple Regression Analysis

Linear regression is a relatively common forecasting technique that employs the Least Square Method (LSM) to find the best fit to the entire data set while minimising the forecasting error. For this study the relationship between different selected indicators can be presented as a set of Linear Multiple Regression(LMREG) equations. For example, any of the collected indicators can be regressed against itself for the previous years (time-lagged). For example the $WCY_{i,t}$ index for the i^{th} economy at time t can be expressed as a non-linear function f below:

$$WCY_{i,t} = f\left(WCY_{i,t-1}, WCY_{i,t-2}, \ldots, WCY_{i,t-n}\right) \tag{5.18}$$

this is expressed in a linear regression format as the following expression:

$$WCY_{i,t} = \alpha + \beta_1 WCY_{i,t-1} + \beta_2 WCY_{i,t-2}, \ldots \beta_n WCY_{i,t-n} + \varepsilon \tag{5.19}$$

where α is a constant, $\beta_1,\ \beta_2, \ldots, \beta_n$ are the regression coefficients, i denotes a certain economy and t signifies the present year, $t-1$ is the previous year, $t-2$ is two years ago and so on. The random error of the series is represented by ε and this error can create a compound random error from the accumulations of previous errors.

Similarly, any of the indicators can be regressed against the other indicators and for the previous years as well, for example:

$$WCY_{i,t} = f\left(KEI_{i,t}, GCI_{i,t}, NRI_{i,t}, GII_{i,t}, IDI_{i,t}\right) \tag{5.20}$$

the above expression in a linear regression format is expressed as:

$$WCY_{i,t} = \alpha + \beta_1 KEI_{i,t} + \beta_2 GCI_{i,t} + \beta_3 NRI_{i,t} + \beta_4 GII_{i,t} + \beta_5 IDI_{i,t} + \varepsilon \tag{5.21}$$

$WCY_{i,t}$ is expressed as a linear function of other indices for the ith economy at time t. If the goal is to track a specific KBE, one can utilise the above equations to predict the next value in the series which can be substituted into the equation to make further future predictions. The full factorial regression models for this study would be a combination of 60 different models.

5.9.2 Panel Data Analysis: Time-Series Cross-Sectional

Panel data analysis is considered to be the flagship of advance econometrics models and it is a technique of studying a specific entity (cross-sections) over a particular time frame (time -series). Hence, it is famously known as the Time-Series Cross-Sectional (TSCS). With over time or repeated observations of enough cross-sections, panel analysis permits study of the dynamics of change with short time series (Yaffee, 2003). A panel data set consists of two parts; a cross section such as (countries, states, districts, firms, economies, or individuals) and a time series, with data gathered on the same individuals, firms or countries for each time period. Panel data analysis can deal with both parts simultaneously, which allows for rich and powerful study of a set of entities (Girma, 2008).

Panel Data Set Structures

Panel data sets generally include cross-sections of data for each individual entity tracked over time periods. The main difference between a time series and a panel data set is that, with respect to the time series, data is collected on a single entity over a long period of time, while in the panel data set the observations are on many entities but at relatively few times - almost always four or less (Markus, 1979). The compiled data set for this study fits the descriptions of the typical case of panel data, where the number of economies is much larger than the number of time periods and this is referred to as a "short panel" data set. In addition because the scores for each economy are spotted every year, it is called a balanced panel (Baltagi, 2005). Table 5.3 shows a portion of the collected data to illustrate the difference between the balanced and unbalanced panel: In the left side of the table, two countries (1 and 2) are observed over three years (2009, 2010, and 2011). Because each economy score is reported every year, the left side set is called a balanced panel, whereas the data set on the right side is called unbalanced panel, because, economy 1, for example, was not observed in year 2011 and economy 2 score is not reported by WEF index in 2010, etc. In general a TSCS regression can be expressed as follows:

$$
\begin{aligned}
Y_{it} &= \alpha_{ik} + \beta' x_{itk} + \varepsilon_{it} \qquad (5.22) \\
& \quad i = 1, \ldots, N; t = 1, \ldots, T; k = 1, \ldots, K
\end{aligned}
$$

where Y_{it} is the dependant variable (to be predicted), x_{itk} is the independent variable, β' is $K \times 1$ and x_{itk} is the i^{th} observation on K explanatory variables and in this case it is the knowledge indicators. The subscript i denotes entities, and it represent the cross-section dimension hence (an economy): $i = 1, \ldots, N$; t signifies time, and it represents the time-series dimension, hence: $t = 1, \ldots, T$ and ϵ_{it} is the random error term.

Tracking the progress of multiple economies over time fits the descriptions of a panel data set, therefore this study suggests using the TSCS regression, which differs from a regular time-series, cross-section, simple or multiple regression in that it has a double subscript on its variables, to symbolise the capture of relationships between the different entities' economies and to take advantage of the behaviour of these economies over time.

For more on panel data analysis and the other different panel data models listed here, the reader is advised to consult the following publications: (Baltagi and Li, 1992; Hsiao, 2003; Yaffee, 2003; Baltagi, 2005; Maddala, 2009).

Table 5.3: Sample data, illustrating balanced and unbalanced panel data set.

Balanced Panel Data							**Unbalanced Panel Data**					
Eco	Year	IDI	KEI	GCI	WCY		Eco	Year	IDI	KEI	GCI	WCY
1	2009	4.3	4.0	1.2	1.8		1	2009	4.3	4.0	1.2	1.8
1	2010	4.4	3.7	1.4	1.3		1	2010	4.4	3.7	1.4	1.3
1	2011	4.3	3.9	2.0	0.7		2	2009	8.4	9.4	7.5	7.4
2	2009	8.4	9.4	7.5	7.4		2	2010	8.7	9.2	N/A	7.6
2	2010	8.7	9.2	7.5	7.6		2	2011	8.4	9.1	7.9	8.2
2	2011	8.4	9.1	7.9	8.2		3	2009	8.0	N/A	7.8	7.7

5.9.3 Computational Intelligence Forecasting Methods

Many researchers have introduced various CI techniques including ANN and SOM models to forecast with complex, non-linear, short time series or missing data.

5.9.3.1 Artificial Neural Network Techniques

Artificial Neural Networks are becoming the trend for their precision in predictions, clustering, modelling and trend analysis. Some techniques are more popular than others, and the ANN are now considered to be the most popular fitting tool with high predictive accuracy compared with other CI methods and of course compared to the traditional linear statistical methods. In many studies, it has been shown that ANN can model any functional linear and non-linear relationship, and that such models are better than regression, since regression is essentially a linear technique used to solve non-linear problems. However, building a neural network for a certain forecasting problem is not an easy task, because many parameters must be considered to achieve the best performance like the number of layers, the number of hidden neurons in each layer, the activation function, the training method, data normalisation etc. (Wilson et al., 2002).

ANNs are commonly categorized in terms of their corresponding training algorithms; mainly supervised and unsupervised training. Back propagation is a common method of training feed-forward ANNs. A feed-forward neural network is usually used for applications that require fitting a set of inputs to a particular targeted outputs (Wilson et al., 2002). Figure 5.4 shows a basic structure of a backpropagation ANN which is usually used for applications that require fitting a set of inputs to a particular targeted outputs. Training this type of network usually happens in three steps: each input $x_1, ..., x_n$ will be fed-forward to train the network to capture the data pattern then it sends the signal of this pattern to the hidden neurons. The hidden neurons $1, ..., h$ compute the activation function using either the binary sigmoid function (0, 1) or the bipolar sigmoid function (-1, 1), and send the results to the output Y_h, by using the gradient descent method, the error generated will be back-propagated after minimising the sum squared error of the outputs against the specified targets. The network keeps cycling through the entire set of training vectors (each complete cycle is called an epoch)

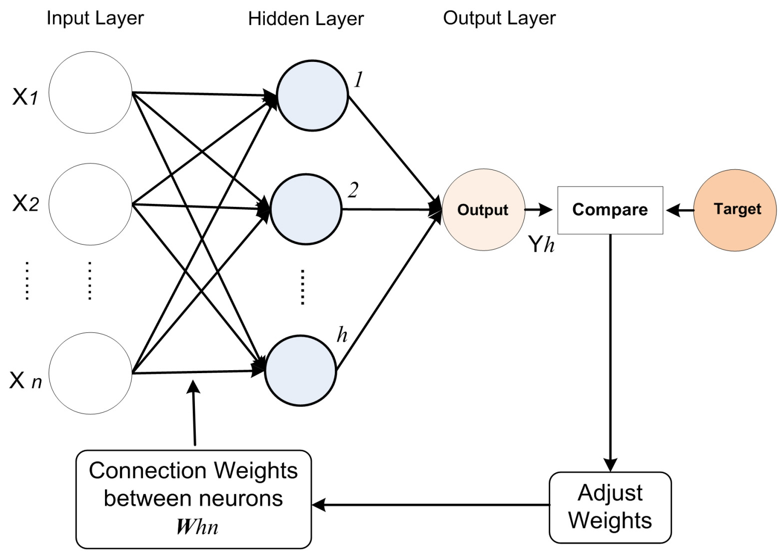


Figure 5.4: Standard backpropagation ANN model.

and keeps adjusting and updating the connection weights W_{hn} accordingly, till it reaches the least possible error or what is known as the global minima (Fausett, 1994).

5.9.3.2 Self-Organising Map

A Self-Organizing Map (SOM) or Self-Organizing Feature Map (SOFM) is a type of ANN, developed by Kohonen Kohonen (1990), and is trained using unsupervised learning to produce a low-dimensional, discretized representation of the input space of the training samples, called a map. The map is an array of $m \times n$ processing neurons. Figure 5.5 shows a basic 4×4 SOM map with any number of inputs variables. Training SOM is totally data-driven and almost no information about the input data is required (Merlin et al., 2010). SOM learns to classify input nodes according to how they are grouped in the input space, therefore, it could recognise countries according to their reported score and it would organise those with similar scores and show them as neighbours even if they are geographically not. Figure 5.6 illustrate the before and after training for a set of random

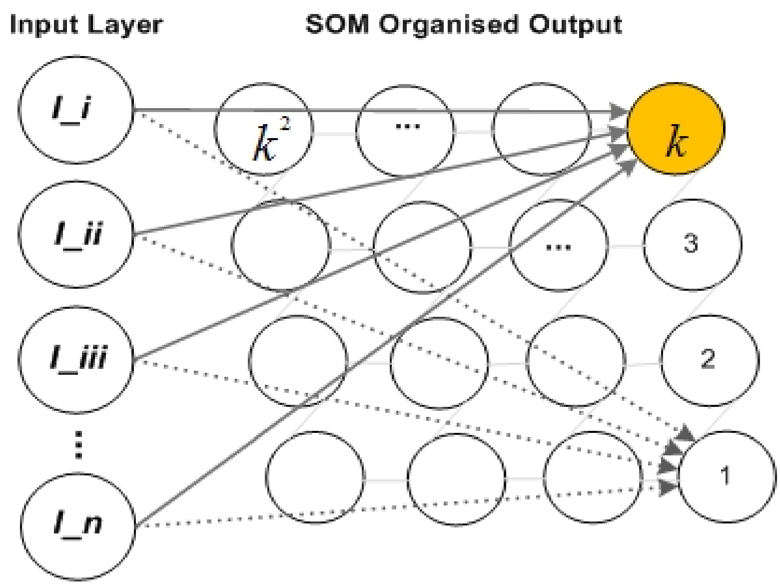


Figure 5.5: Standard SOM model.

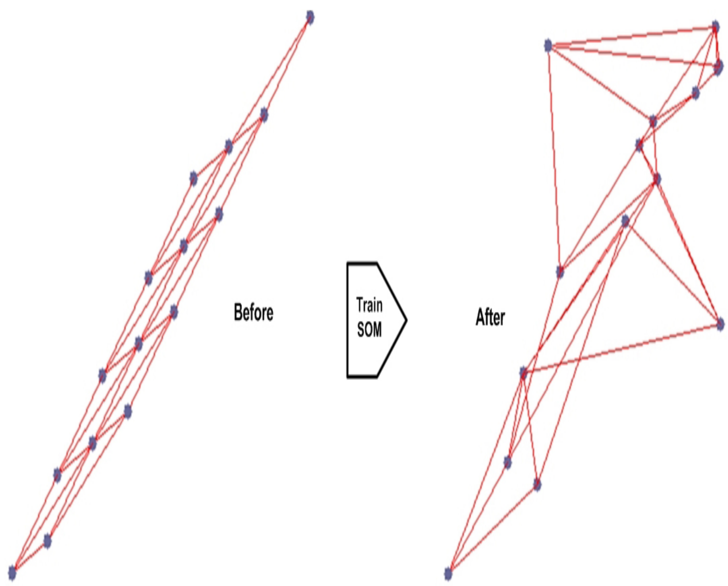


Figure 5.6: SOM neuron weight positions of random data points before and after training for one-step ahead clustering map.

variables. Thus, this shows that self-organizing maps learn both the distribution and topology of the input nodes they are trained on. Also, it is capable of providing a visual, easy to interpret, distribution-free and non-linear description of the multidimensional data distribution without losing the topological relationships of data and sight of individual indicators (Sarlin, 2011).

SOM is also a suitable technique for multidimensional data clustering (Ilonen et al., 2006) and it helps to avoid the curse of dimensionality and the heterogeneity bias which arise among cross-sectional units (Hsiao, 2003). Thus, the aim is to find similar countries (not necessarily territorial neighbours) by using comparative measures and sensitivity analysis. This can be achieved arithmetically by computing the first and the second derivative for each economy along the time series (Zimmermann, 2011). Alternatively this can be achieved manually by grouping similar countries based on their similar scores and observing the moving ones. The task becomes daunting when presented with many countries, many reported scores with high variations between their reported scores. If they are similar and should be classified as neighbours. An automated method to find similar competitive KBE within multi-dimensional data set would be highly desirable.

Although SOM algorithm is a well established non-linear mapping tool, and it has many beneficial properties, such as tolerance for incomplete and small data set (Ghaseminezhad and Karami, 2011), however, a few issues need to be tackled to get efficient results. It is suggested by Thang et al. (2003) to train the SOM in two phases: ordering phase and then tuning phase. The ordering phase helps the network to quickly scan a large area in search of related neurons, and not getting stuck in a local minima. The ordering phase usually requires setting high learning rate, large distance and small number of epochs. The tuning phase requires higher number of epochs, small distance and low learning rate to tune-up the rough structure of the earlier phase to produce a well organised and tightly coupled map. It is also suggested to use some heuristics measures to evaluate the efficiency of the trained SOM by measuring and comparing the Quantization Error (QE) and the Topographic Error (TE), eventually aiming for a low TE and QE.

5.10 Robustness and Validation Analysis

The applicability of any newly created model depends on its validity, stability, and soundness. Robustness analysis plays a crucial part in the after development stage of any developed system. The model goes through an "X-ray" testing to determine the coherency between the system inputs and outputs (Kennedy, 2003).

5.10.1 Monte Carlo Analysis

The robustness of a composite indicator can be tested by subjecting the indicator to simulated environments that are based on the different formulae that make the underlying model of the indicator. One of the most widely used methods for evaluating a system's robustness is the Monte Carlo approach , which is based on multiple evaluations of the model with k randomly selected model input factors. The procedure has four steps:

1. Assign a domain of possible input factors.
2. Generate random inputs from a probability distribution of independent input factors.
3. Run the simulation computation on the input factors and return the mean estimation of output vectors.
4. Analyse the results of output vectors (Saltelli et al., 2000).

5.10.2 Adaptive Neuro-Fuzzy Inference System

Creating a brand new composite indicator is not only challenging but also needs validation of the accuracy of the new product. The validation of a new SCI can be done through an in-line or permutation predictions of the combinations of the different variables making the new SCI. Using statistical methods to predict the variables from each other might produce inaccurate result and would be a repetitive process especially if the number of variables making the new index are high. CI methods such as the Adaptive Neuro-Fuzzy Inference System (ANFIS) (Jang, 1993), can be a tremendous help as they known to learn to produce an

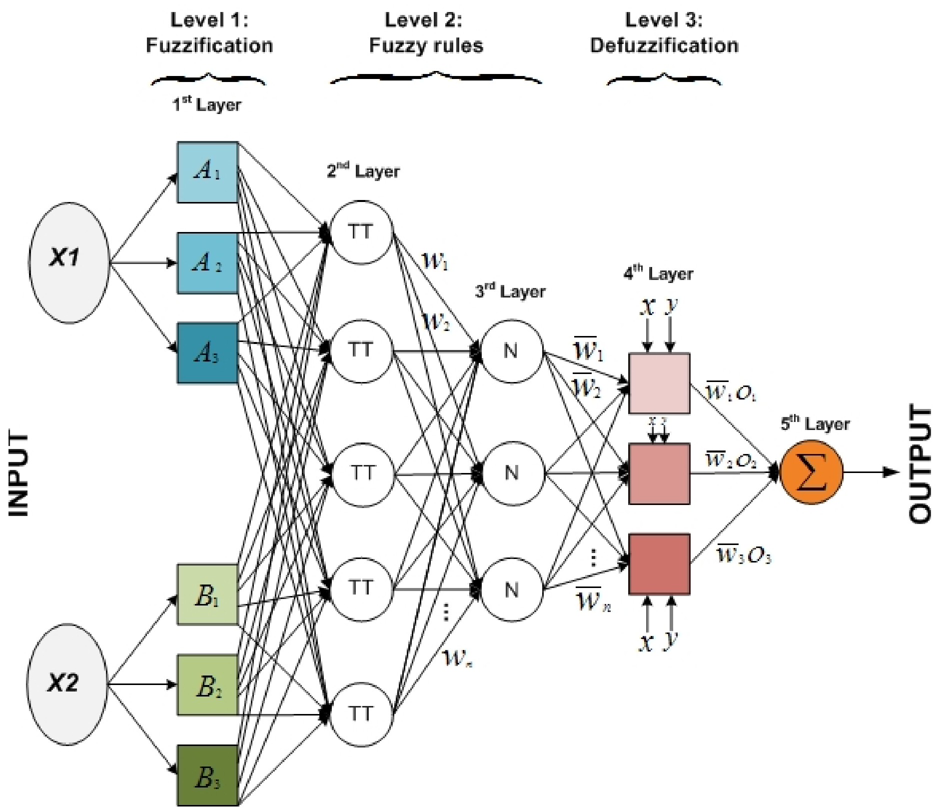


Figure 5.7: ANFIS structure with two inputs for one targeted output.

accurate predictive results. ANFIS are fuzzy models put in the framework of adaptive systems to facilitate learning and adaptation. By making use of both the neural networks adaptability and the fuzzy qualitative method, it can be trained to map precisely non-linear functions and predict a chaotic time series with or without expert(s) knowledge, and/ or specified source-target data sets.

Figure 5.7 shows the ANFIS structure (Hopgood, 2011), where a first order of Sugeno fuzzy model is used as a means of modelling fuzzy rules into a desired targeted output. In the depicted diagram the square represents an adaptive node (the parameters are changed during training), while a circle represents a fixed node. This structure consists of five layers of feed-forward neural topology. The nodes functionality in each layer can be summed as follows: The first layer is adaptive and consists of neurons of linguistics labels. The output of this layer can be presented as a membership function:

$$L1_i = \mu A_i(x_i) \tag{5.23}$$

The second layer consists of fixed nodes. The purpose of this layers is to estimate a rule firing strength (wi), which is calculated as the multiplication of the received signal:

$$L2_i = \bar{w}_i = \mu A_i(x_i)\mu B_i(x_n) \tag{5.24}$$

The third layer also consists of fixed nodes. Each node in this layer calculates the ratio (wi) of the ith rule's and the total of all the rules firing strengths, represented by j. The output can be calculated by:

$$L3_i = \bar{w}_i = \frac{w_i}{\sum_{j=1}^{i} w_i} \tag{5.25}$$

The nodes in the fourth layer are adaptive and acts as defuzzifier. Each node output can be calculated as the product of the previously calculated relative firing strength of the ith rule:

$$L4_i = \bar{w}_i f_i = \bar{w}_i(p_i x_1 + q_i x_n + r_i) \tag{5.26}$$

The five layer consists of one node to sum-up all the incoming signals from the previous layer as follows:

$$L5_i = \sum_{i=1}^{j} \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{5.27}$$

The final results are defuzzified using the weighted-average. The mission of this structure is to tune all the modifiable parameters so ANFIS output can match the training data (Hopgood, 2011; Keles et al., 2008).

5.11 Summary

This chapter described all the methods and techniques to be utilised throughout this thesis. The chapter started with a brief descriptions of the datasets which was collected and arranged based on the earlier qualitative taxonomy. The procedures and techniques usually used for developing composite indicators were discussed. This includes, but not limited to, data treatments and analysis such as multivariate analysis, correlations, normalisation, missing data imputations, outliers' detection, weighting, aggregation and robustness analysis. This chapter also included alternative CI techniques, FCM specifically was introduced to impute missing data, weight and aggregate variables. This chapter also provided several forecasting techniques. From the CI side, both supervised and unsupervised ANN techniques were considered. From the advance econometrics forecasting models Panel data analysis specifically the TSCS were detailed as a method of choice that fits the descriptions of the collected datasets. In the subsequent chapters will refer back to these methods as they are used within the context and to produce the results. Finally, the chapter concluded with validation and robustness techniques such as Monte Carlo Simulation and the ANFIS from the CI side.

Chapter 6

Unified Macro-Knowledge Competitiveness Framework

6.1 Introduction

In Chapter 4 a Fuzzy Proximity Knowledge Mining (FPKM) model has been introduced to form the qualitative taxonomy for the suggested *UKCI* with little human or "experts" intervention. The taxonomy confirmed to the fact that, well-chosen individual variables are very important to the making of the synthetic index, however, it is not the intention of this study to "reinvent the wheel" but rather to work with the existing KBE indicators to combine their strengths and avoid all weaknesses when constructing the intended macro-knowledge framework. Hence, the produced taxonomy was able to collect the main indicators that would measure the KBE progress and competitiveness; it also "learn" to combine similar sub constituent's elements to make the proposed *UKCI* major units of baskets and sub baskets to serve as a top-down macro- knowledge progress monitoring tools.

To this end this study will present the continuation of the proposed epistemology, qualitative taxonomy, data collection, and the development methods to present the proposed model. It will emerge as a result of different experiments with different solutions in order to build a robust macro-knowledge competitiveness index in accordance with the devised qualitative taxonomy for building the suggested *UKCI*. As a result a new framework to construct a SCI is to be proposed in the process.

In addition to understanding and critically evaluating the existing KBE global indicators, this research main aim is to introduce a new way of developing a new types of SCIs that would overcome the subjectivity of “experts” and the shortfalls of the statistically based SCIs by using computational intelligence means. The proposed intelligent indicators would carry special character as they are data-driven and therefore such indicators have the capabilities to accurately rank nations based on non-biassed “learning”. These indicators are branded as The Intelligent Synthetic Composite Indicators (*iSCI*) hereinafter. *iSCI* reflects a novel vision towards a new breed of SCIs.

As a case study the results presented in this chapter are based on the Information and Communication Technology (ICT) and E-Services real and related variables as it was illustrated in Chapter 4. The process involves using Fuzzy c-Means (FCM) clustering to identify natural aggregation in data from the qualitative taxonomy grouped variables data sets to allow for concise representation of the relationships embedded within the variables and to generate the final *UKCI* composite indicator. Different methods are investigated for the purpose of missing data, weighting, aggregations of variables into a smaller subset while avoiding any organisations or experts subjectivities or opinions biases. FCM and its derived strategies, specifically, the Optimal Completion Strategy (OCS) and the Nearest Prototype Strategy (NPS) are also investigated for missing variables scores.

The layout of this chapter is organized as follows: Section 6.2 includes the proposed framework and the steps taken to develop the new *iSCI* is explained. Section 6.3 presents an empirical case study with comparative efficiency, limitations and results for the different utilised methods followed by robustness analysis of the constructed intelligent composite indicator. In Section 6.4 the (*iSCI*) model is generalised to create the suggested *UKCI*. Section 6.5 apply the *UKCI* on the MENA region, to assess its applicability, effectiveness and added values. In Section 6.6 the *UKCI* final scores and ranks results are listed. The chapter summary is provided in Section 6.7.

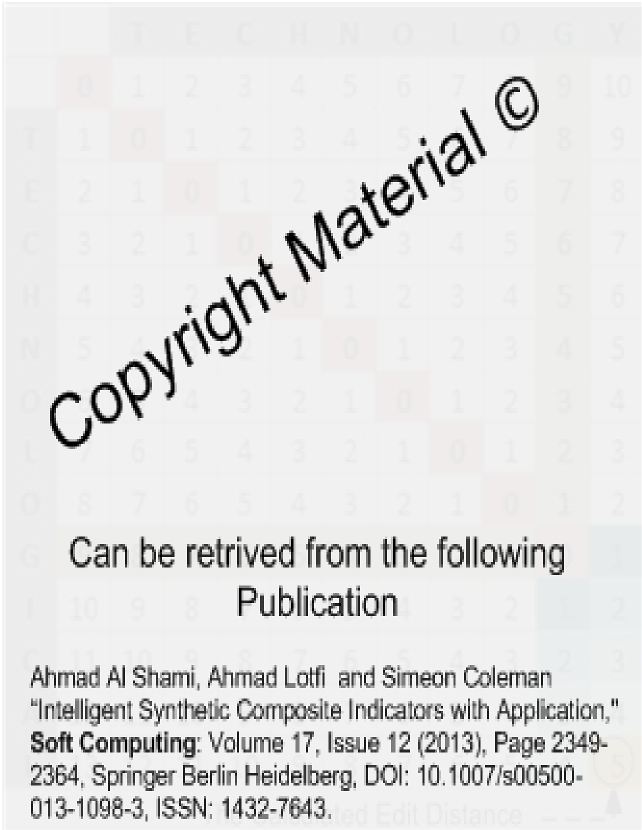


Figure 6.1: Schematic diagram of the proposed intelligent indicators development.

6.2 Intelligent Synthetic Composite Indicators Framework

To overcome the shortfall of the SCI, a new framework to generate the *iSCI* is proposed. Figure 6.1 shows a schematic diagram of the proposed framework, which comprises five stages:

In the first stage, only the qualitative information will be dealt with. This include the construction of qualitative taxonomy or theoretical framework for the proposed index. In this stage, a web crawler is used to mine the web content of certain targeted keywords referred to as seed URLs to gather the needed variables. Then the definitions and variable names used by all the gathered indicators are analysed using FPKM technique. They are mainly text based information and the outcome of this stage should produce the desired taxonomy for the new index.

In the second stage of the proposed framework, the numerical or quantitative information gathered from the first stage will be collected and analysed. This includes correlation, outliers detection and multivariate analysis. In addition, in this stage, missing information will be imputed comparing different techniques. In the third stage of development, all numerical values are standardised to a uniform unit of measurements. Min-Max normalisation technique is used to transform different scores ranges to a value between 0 and 1. In the fourth stage, different weighting and aggregation techniques are compared. Statistical methods are compared against FCM to aggregate the indicators into a smaller subset while avoiding the curse of dimensionality in the data, and also avoid any organisations or experts subjectivities or opinions biases. Finally in the fifth stage, the validity and robustness of the proposed framework is tested.

6.3 Empirical Case Study

In this section, a set of empirical tests on a real case study is presented. The case study produces a new unified ICT index, to illustrate the methods of construction for *iSCI*, and to compare the effectiveness of the different methodologies suggested so far.

6.3.1 Numerical Dataset Analysis

The dataset used in this case study is collected based on the qualitative taxonomy for the Unified ICT baskets. The ICT variables to be used here are presented in Table 4.1. These datasets are freely and readily available from the annual reports issued by the organizations mentioned earlier. The set contains various number of economies as reported by the sources, which was each given a score and a rank across the eleven filtered variables for three consecutive years 2009, 2010 and 2011. In the subsequent sections a couple of measures are taken to test to what degree the ICT variables are related and comparable.

6.3.1.1 Correlation

To test the degree of relationship between the filtered variables correlation analysis is conducted initially on the collected data set. The correlation coefficient matrix between the ICT related variables is summarised in Table 6.1.[1] The result revealed a "moderate" to "strong" positive relation between almost all indicators. The highest correlation coefficient value is 0.967, which is significant and occurred between the WEF-NRI, Individual Usage - coded as I - and the ITU-IDI, ICT Use - coded as K. The lowest correlation is 0.379 "weak" which resulted between the WEF-NRI, Government Readiness - coded as G - and ITU-IDI, ICT Price - coded as H. Based on these results we can settle that the indicators are correlated and comparable. Several studies including Roessner et al. (1996), Porter et al. (2009), and Johnson et al. (2010) investigated the relation of high technology competitiveness indicators and concluded a similar result that these indicators complement each other, and their differences are mainly due to the limitations and variations of the traditional methods used to weight and aggregate the input variables. Hence, it would be highly desirable to unify such efforts for a full rounding result. Therefore, it is feasible to normalise, and aggregate the efforts of these ICT indicators into a "one for all" solution that would reflect, measure and rank the combined level of ICT and e-services in and between countries.

6.3.1.2 Outlier Detection

To check for any outliers within the collected dataset, the Mahalanobis Distance is used (Maesschalck et al., 2000). It is a distance measure based on correlations between variables to detect any point that has a greater distance from the rest of the sample. The result of Mahalanobis distance test as depicted in Figure 6.2, spotted two points as they are slightly far from the rest of the countries. The

[1] The rest of this text is censored as a Copyright Material which can be retrieved from the following article: Ahmad Al Shami, Ahmad Lotfi and Simeon Coleman "Intelligent Synthetic Composite Indicators with Application," Soft Computing: Volume 17, Issue 12(2013), Page 2349-2364, Springer Berlin Heidelberg, DOI: 10.1007/s00500-013-1098-3, ISSN: 1432-7643.

Table 6.1: Correlation coefficient matrix for individual variables.

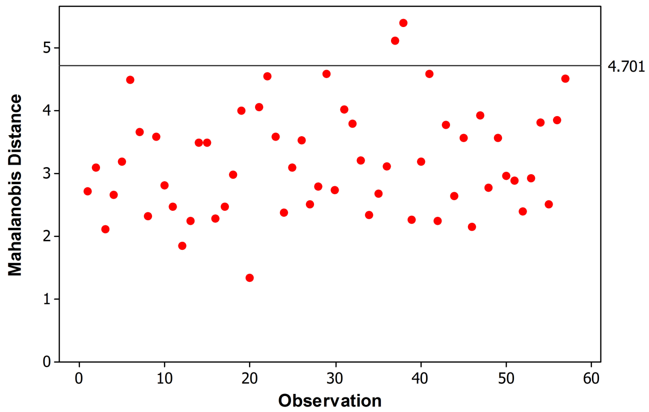


Figure 6.2: Outliers detection between variables, N=57.

Existence of such outliers is not problematic, therefore it is decided to keep them in the dataset.

6.3.1.3 Multivariate Analysis

PCA is one of the multivariate and inputs reduction methods. The goal of PCA is to reveal how different variables change in relation to each other and how they are associated. PCA is useful when there are two or more variables, and believe that there is some redundancy in those variables. In this case, redundancy means that some of the variables are correlated with one another, possibly because they are measuring the same construct. Because of this redundancy, it should be reasonable to reduce the observed variables into a smaller number of principal components "artificial variables" without significant loss of information. PCA is employed in this study to serve three purposes: first, to test if the eleven variables could be reduced. Second, to reduce the number of indicators to a smaller subset. Third, to forsee the possibility of filtering out the trivial components before is used. The trivial components usually act as noise and could stand in the way of getting a sound and meaningful clustering result. Figure 6.3 shows the result of the PCA analysis: the first component display the highest eigenvalues as it explain 76.25% of the variability in the data, the second 9.42%, the third component accounted for 4.52% and so on. The results of the scree test suggest that only the first two components are meaningful. Therefore, only the first two components

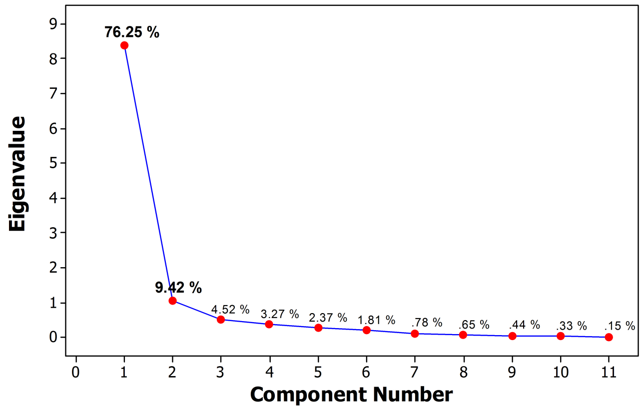


Figure 6.3: PCA result showing the scree plot of eigenvalues of covariance for 57 countries.

were retained. Combined, components 1 and 2 accounted for 85.67% of the variability in the data, which we can retain. The plot levels off after the second component where the rest of eigenvalues that represents the trivial components of 14.33% which can be discard.

6.3.1.4 Variables Standardisation

Normalisation usually is used to transform different measurement units into a uniform unit, so they can form a clear comparable elements, and to avoid problems in mixing measurement units (e.g. money, talent, skills) (Freudenberg, 2003). The issue at hand is not the use of different measurement units, but the scores scale ranges. Hence, to unify the score ranges between the different selected indicators, Min-Max normalisation -as formulated in Equations 5.3- was applied by taking all the different scores ranges collected in the data set and transforming these to a value between 0 and 1, where the lowest (min) value is set to 0 and the highest (max) value is set to 1. In the cases where a high value implies inferior result such as ICT Price, we resort to the reverse Min-Max normalization process as in Equation 5.4, so that, in addition to converting the series into a $[0 - 1]$ range, inverts it, so that 0 implies poor and all the way to 1 as the top possible performances.

6.3.2 Missing Data Imputation

By studying the nature of the collected indicators datasets from various sources, one can clearly notice that the data is usually present for developed economies, but it is the opposite when it comes to underdeveloped or developing nations. Such cases force the perpetrators of these indicators to substitute the missing data, estimate, or drop the country completely from the ranking list, if most data figures are unattainable. By also examining the collected data set and after correlating the indices over themselves for three consecutive years 2009, 2010 and 2011, a very high degree of correlation can be noticed. This signifies that development and progress in the technology sector are a slow processes as suggested by the slight increase/ decrease or constant scores value. This indicates a need for an accurate technique to capture the slight and unnoticed differences to substitute for the missing values.

This research study applied the “in sample (train)-out of sample (test)” logic in order to test for a suitable imputation method (Kumiega and Vliet, 2008; Jammazi and Aloui, 2012). This logic consists of taking the complete part of the dataset, leave some of the data out of the sample(for the same countries and in the same proportion of the complete dataset). Train the imputation model using the remaining data “in sample”. Then test the accuracy of the applied imputation methods using the “out of sample” data.

The “in sample-out of sample” is applied as follows; first, countries “observations” that contain any missing values were removed, to end up with a complete portion of the dataset. This step produced a dataset consisting of fifty-seven countries, for three consecutive years 2009, 2010 and 2011, for the eleven ICT and e-services related indicators. Second, a random choice of fifteen countries is made, followed by an artificial censoring of the data (in increments of 10% up to 50%) from the complete set. Third, training each of the suggested imputation classifier on the remaining portion of the data set. Fourth, the accuracy of each imputation model is checked using the “out of sample” data set.

The effectiveness of the employed methods were checked using two accuracy methods: Index of Agreement, d by Willmott et al. (1985), where a model prediction error varies between 0 and 1; $d = 1$ indicates a perfect match, and $d = 0$

indicates no agreement. d is defined as:

$$d = 1 - \left[\frac{\sum_{i=1}^{N} (I_i - O_i)^k}{\sum_{i=1}^{N} \left[\left(|I_i - \overline{O}| + |O_i - \overline{O}| \right)^2 \right]} \right]^2 \tag{6.1}$$

where N is the number of imputations, I_i is the value of the imputed data point, O_i is the original data points and $\overline{O}$ is the average of the original data.

The Mean Absolute Error (MAE) is used as another accuracy measure. The smaller the MAE values, the better the imputation method. It is also expressed as follow:

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |I_i - O_i| \tag{6.2}$$

6.3.3 Missing Data Accuracy Comparative Results

The comparative results of the different imputations methods used in this study and their accuracy measures, using MAE and the index of agreement, d, are presented in Tables 6.2 and 6.3 respectively.[1]

After ten trials run for the imputation methods, all methods tested have produced very good results with a slight error of classifications for each level of missing values. For missing values of (10%, 20% and 30%), OCS outperformed the rest in both accuracy measures MAE and d, where higher accuracy are presented in bold. But when it comes to higher percentage of 'missingness' i.e., 40% and 50%, the regression method performed slightly better. For overall performance with a certain degree of confidence (95%); a confidence interval plot is presented

[1]The rest of this text is censored as a Copyright Material which can be retrieved from the following article: Ahmad Al Shami, Ahmad Lotfi and Simeon Coleman "Intelligent Synthetic Composite Indicators with Application," Soft Computing: Volume 17, Issue 12(2013), Page 2349-2364, Springer Berlin Heidelberg, DOI: 10.1007/s00500-013-1098-3, ISSN: 1432-7643.

Table 6.2: Mean absolute error of 10 trials for the tested imputation methods.

Table 6.3: Index of agreement of 10 trials for the tested imputation methods.

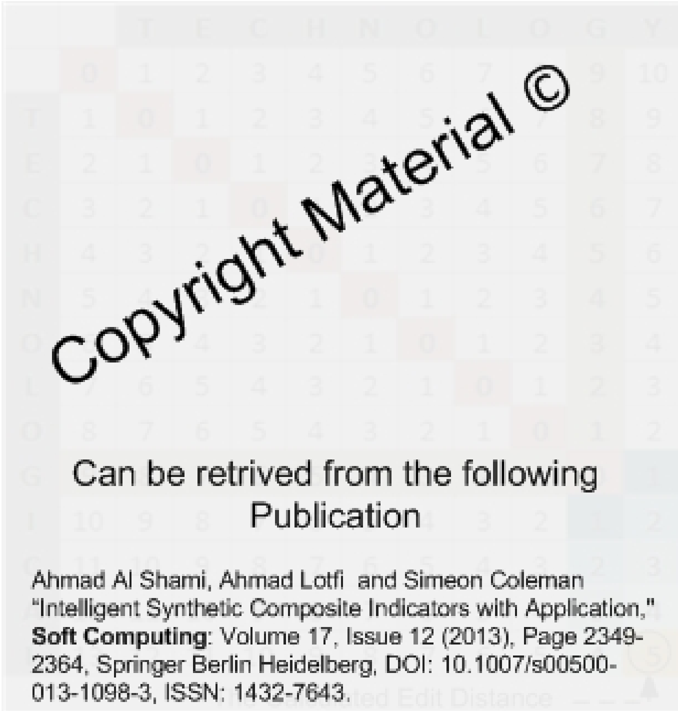


Figure 6.4: Imputation methods MAE overall performance, 95% confidence interval.

in Figure 6.4, which shows that even though the FCM based method, OCS and NPS, have wide intervals, the average performance is as good and competitive as the regression or even the expectation maximisation methods. This means that for application and data set, it is feasible to substitute missing values using any of the presented computational based techniques presented here, which also could have other added values such as simplicity, speed and convergence.

6.3.4 Weighting and Aggregation Schemes

In general a weight given to a certain variable can highly influence the outcome of SCI and the overall country rankings. There are many traditional statistical weighting techniques, for example, equal weighting, data envelopment analysis, factorial analysis, and benefit of the doubt. There are also experts and/ or stakeholders participatory methods such as analytic hierarchy processes, conjoint analysis and budget allocation processes. Similarly, the arithmetic mean, geometric mean, additive rules etc. are regularly used for the purpose of aggregating the variables to form a single value, hence a "composite index" (OECD, 2008a). On other hand, computational intelligence techniques such as ANN, SOM, fuzzy systems and recently some hybrid methods such as hesitant fuzzy geometric means and intuitionistic fuzzy hybrid geometric operators have been proposed and applied to act as an aggregator for multi input - single output systems. Such methods were made to help decision makers to effectively deal with multiple attribute decision making under hesitant or intuitionistic environments (Zhu et al., 2012) and (Zhao and Wei, 2013).

6.3.4.1 FCM Weighting and Aggregation

For the purpose of this study, it is suggested to use the FCM algorithm as stated in Section 5.7.1, to cluster available data for each country and identify a cluster centre, which is weighted by the mean of all points based on their degree of belonging to the cluster, to represent the ICT index score for that economy. Each score point will have a degree of membership to the scores clusters, rather than belonging entirely to just one cluster. Therefore, scores on the border of a cluster could be in the cluster to a smaller degree than points in the middle of cluster.

The degree of membership of belonging to a certain cluster, is related to the inverse of the distance to the cluster centre which can be stated using Equation 5.16, where the centroid of a cluster v_i is the mean of all points, weighted by their degree of belonging to the cluster.

Different fuzzy clusters can be characterised by a class membership function matrix, and cluster centres are determined first at the learning stage, and then the classification is made by the comparison of euclidean distance between the incoming features and each cluster centre. To better visualise the achieved result and to consider the limitation of the dataset (data records) and rather large features (economies), only two clusters are formed. These two clusters are representing a range where the predicted value belongs. Figure 6.5 presents a sample for two randomly selected economies showing the fuzzy cluster centres versus the average scores. It is clear that the fuzzy cluster centres are located within the entropy of

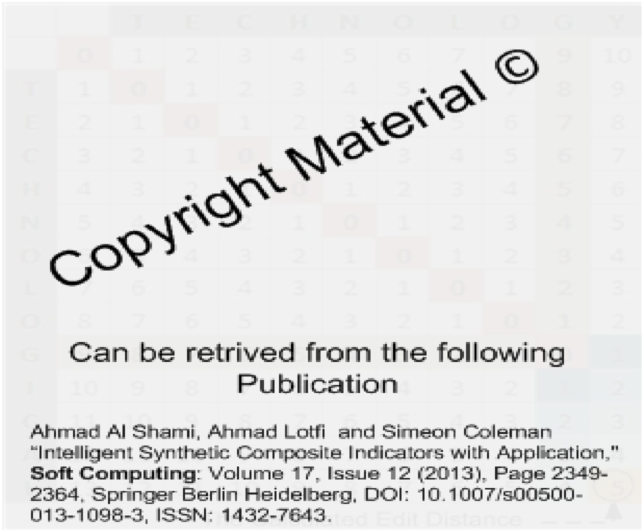


Figure 6.5: Illustration of the fuzzy clusters for two selected economies.

the input variables and hence, would represent the optimum value for the sampled economies.

6.3.4.2 LBG VQ Weighting and Aggregation

The LBG VQ algorithm, is similar to the classical Vector Quantization (VQ) technique in that it generate both the codevectors and codebook. The LBG VQ is an iterative segmentation method that requires an initial codebook C. The average of the full training sequence is initially used as the start codevector value. This value is then split to two halves, next the algorithm uses these two codevectors as the initial codebook. The two codevectors later are split into four and so on, till the chosen number of codevectors is achieved. The full VQ and LBG VQ process and equations are explained in Section 5.7.2 and Appendix A respectively. Figure 6.6 shows the final score for LBG VQ vs the FCM score for two selected economies. This shows a close similarity between the LBG VQ and FCM obtained final scores which requires further investigation as to what would constitute a more robust final score value.

Table 6.4: Eigenvalues of ICT index dataset.

	Eigenvalue	Vari% Total	Cuml.	Cuml.Vari%Tot

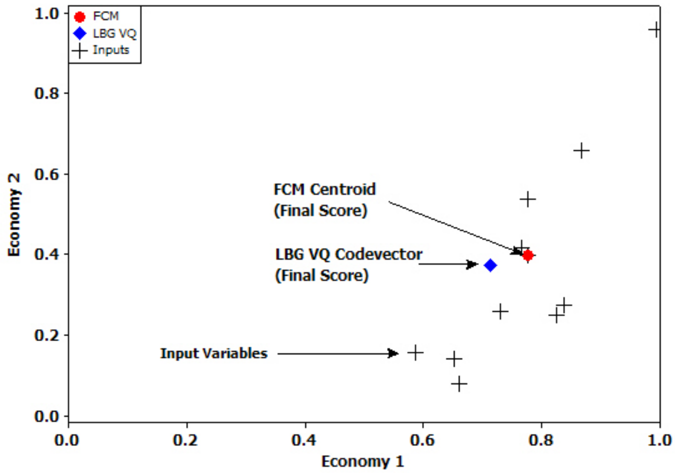


Figure 6.6: Illustration of the FCM vs. LBG VQ final scores for two selected economies.

Table 6.5: Variables weights extraction using PCA & FA rotated factor loadings for ICT index.

6.3.4.3 Statistical Weighting and Aggregation

PCA, and more specifically Factor Analysis groups together individual indicators which are collinear to form a composite indicator that captures as much as possible of the information common to individual indicators. Each factor (usually estimated using PCA) reveals the set of indicators with which it has the strongest association. The idea under PCA and FA is to account for the highest possible variation in the indicator set using the smallest possible number of factors. FA involves several steps: Correlations, factor extraction, rotation of factors and then construction of the weights (Johnson and Wichern, 2007).

Table 6.4 lists the prominent factors and eigenvalues of the eleven individual variables that compose the ICT index. We are interested in [1]

Table 6.5 shows the results for the two factors loadings of ICT index based

[1]The rest of this text is censored as a Copyright Material which can be retrieved from the following article: Ahmad Al Shami, Ahmad Lotfi and Simeon Coleman "Intelligent Synthetic Composite Indicators with Application," Soft Computing: Volume 17, Issue 12(2013), Page 2349-2364, Springer Berlin Heidelberg, DOI: 10.1007/s00500-013-1098-3, ISSN: 1432-7643.

on principle components extractions following rotation and normalisation.[1]

To aggregate the eleven indicators based on the PCA & FA weighting scheme, this study resorted to the Geometric Mean Equation (GME), which is stable and highly recommended for aggregating SCI, and such technique is widely studied in fuzzy set theory (Zimmermann and Zysno, 1983). The GME can be expressed using the following equation:[2]

(6.3)

where SCI is the aggregated composite score for a certain country, c. X is the value of each variable or indicator, i. W is the weight assigned for each variable or indicator. Figure 6.7 shows the final scores for GME vs LBG VQ vs FCM final scores for two selected economies. This illustration shows that GME scores are skewed and far from the rest of the actual input variables compared to the other CI based scores. Further investigation and validation is necessary to clearly proof which of these three suggested final scores would constitute a more valid, robust and non-bias final score values?

In general, there are downside aspects to the use of PCA and FA analysis approach. A general problem with these methodologies is that they are sensitive to modifications in the basic data. Data revisions and updates, possibly implying additional observations (such as the inclusion of new countries), may change the set of weights (i.e. the estimated loadings) that are used to compute the summary indicators. The results are also likely to be sensitive to the presence of outliers, which may introduce a spurious variability in the data, and may as well suffer from small-sample problems, which are particularly relevant when the focus is on a limited set of countries. Finally, data limitations may imply difficulties in the statistical identification and the economic interpretation of the unobserved factors (Nicoletti et al., 1999).

[1] The rest of this text is censored as a Copyright Material which can be retrieved from the following article: Ahmad Al Shami, Ahmad Lotfi and Simeon Coleman "Intelligent Synthetic Composite Indicators with Application," Soft Computing: Volume 17, Issue 12(2013), Page 2349-2364, Springer Berlin Heidelberg, DOI: 10.1007/s00500-013-1098-3, ISSN: 1432-7643.

[2] The rest of this text is censored as a Copyright Material which can be retrieved from the following article: Ahmad Al Shami, Ahmad Lotfi and Simeon Coleman "Intelligent Synthetic Composite Indicators with Application," Soft Computing: Volume 17, Issue 12(2013), Page 2349-2364, Springer Berlin Heidelberg, DOI: 10.1007/s00500-013-1098-3, ISSN: 1432-7643.

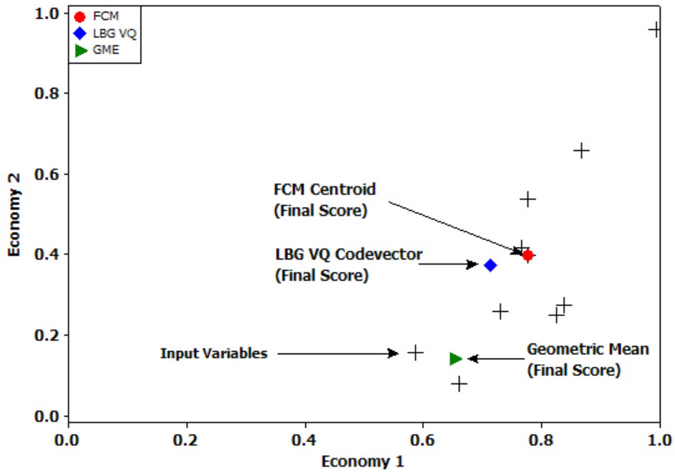
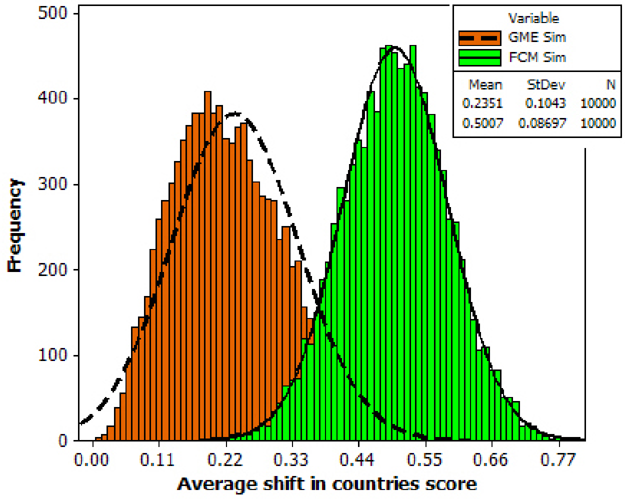


Figure 6.7: Illustration of the GME vs. FCM vs. LBG VQ final scores for two selected economies.

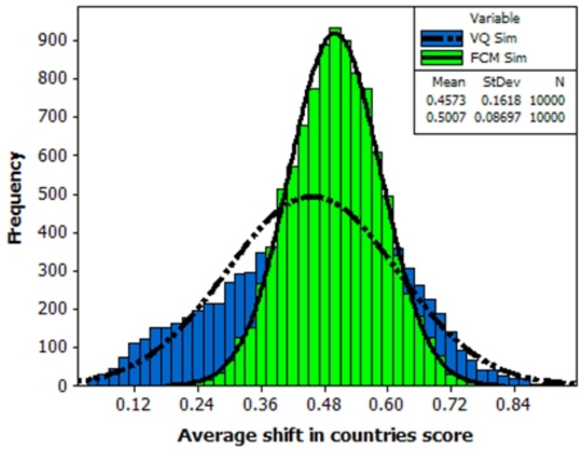
6.3.5 Validation and Robustness Analysis

To compare the degree of uncertainty of the achieved scores and ranks for the PCA - FA & GME combined vs. LBG VQ vs. the FCM aggregation technique, it is instructive to validate the robustness of the new ICT index; a Monte Carlo simulation is conducted as a single experiment with $10,000$ uniform random variables for each of the 11 inputs, then injected into all models. The result of the experiment as depicted in Figures 6.8(a), 6.8(b), 6.8(c) and 6.9, show that the used techniques are valid and they would produce normally distributed results, however, the GME model appears to be skewed to the left and far off from the mean of the real scores. The LBG VQ scores are tipped towards the tails of the curve and slightly skewed to the right. When comparing the standard deviation of the models, the FCM model has a smaller standard deviation than the GME and LBG VQ. The total distribution of the simulated data for FCM matches the behaviour of the scores generated using the real variables, hence, FCM method produced a more robust model than the rest.

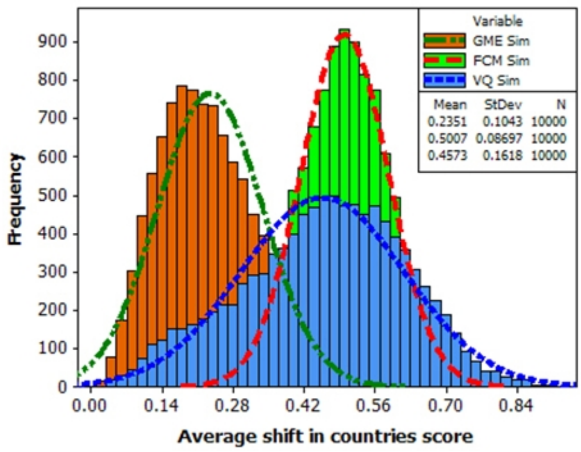
To highlight the dependence of rankings on the different weighting and aggregation methods used (in this case, PCA - FA & GME, LBG VQ versus FCM for the ICT index data set for 2011 with 57 countries), Table 6.6 lists the aggregated scores and ranking. Although the same input variables and the same pre-processing were used, the resulting scores and rankings are different. For example, Sweden ranks first according to FCM and second according to the LBG VQ and PCA - FA & GME weighting and aggregation. While Singapore ranks eighth according to the FCM and tenth with the LBG VQ, but first according to the PCA - FA & GME. The United States, Luxembourg, Canada, Slovak Republic, Malaysia and Argentina suffered the largest shift in their rankings. However, a large number of nations (15 out of 57) experienced stable ranking with only one rank shift between the three different used methods.



(a) FCM vs. GME

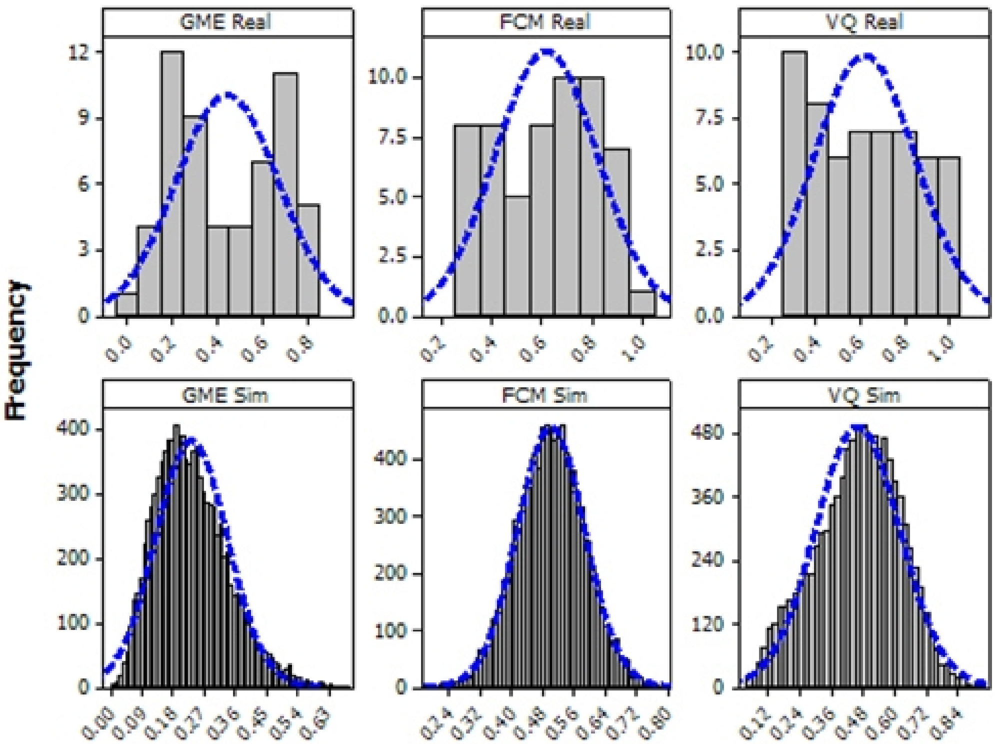


(b) FCM vs. LBG VQ



(c) FCM vs. GME vs. LBG VQ

Figure 6.8: Monte Carlo results for FCM, GME and LBG VQ simulated models.



Real Scores			Simulated Scores		
N= 57	**Mean**	**StDev**	**N= 10000**	**Mean**	**StDev**
GME	0.4444	0.2268	**GME**	0.2351	0.1043
FCM	0.6167	0.2057	**FCM**	**0.5007**	**0.0869**
VQ	0.6186	0.2310	**VQ**	0.4573	0.1618

Figure 6.9: Monte Carlo result of simulated vs. the real data models for FCM vs. GME and LBG VQ.

Table 6.6: Unified ICT index scores and ranks for 57 countries using GME vs. VQ vs. FCM, year 2011.

	PCA-FA & GME		LBG VQ		FCM			PCA-FA & GME		LBG VQ		FCM	
Country	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**	**Country**	**Score**	**Rank**	**Score**	**Rank**	**Score**	**Rank**
Sweden	0.806	**2**	0.998	**2**	1.000	**1**	Bulgaria	0.290	38	0.471	39	0.636	30
Finland	0.723	7	0.867	12	0.943	**2**	Slovak Republic	0.228	45	0.575	32	0.630	31
Luxembourg	0.671	15	0.928	9	0.918	3	Italy	0.343	33	0.527	34	0.618	32
Korea Rep.	0.755	5	0.766	18	0.888	4	Czech Republic	0.391	31	0.643	27	0.616	33
Denmark	0.791	3	0.976	4	0.875	5	Hungary	0.346	32	0.577	31	0.597	34
Iceland	0.696	10	0.978	3	0.859	6	Croatia	0.329	34	0.565	33	0.573	35
Netherlands	0.709	9	0.959	5	0.853	7	Malaysia	0.473	27	0.514	35	0.552	36
Singapore	0.821	**1**	0.903	10	0.853	8	Poland	0.289	39	0.488	38	0.546	37
Switzerland	0.678	13	1.000	**1**	0.845	9	Greece	0.320	35	0.495	37	0.521	38
Norway	0.689	11	0.947	7	0.838	10	Romania	0.249	41	0.386	43	0.503	39
Hong Kong, China	0.774	4	0.954	6	0.828	11	Chile	0.318	36	0.507	36	0.481	40
United Kingdom	0.722	8	0.947	8	0.816	12	Russian Federation	0.263	40	0.362	46	0.481	41
New Zealand	0.625	18	0.710	22	0.796	13	Turkey	0.246	43	0.432	41	0.421	42
Japan	0.608	20	0.700	23	0.791	14	Jordan	0.240	44	0.399	42	0.407	43
Austria	0.631	17	0.783	15	0.781	15	China	0.317	37	0.341	50	0.405	44
Germany	0.680	12	0.833	14	0.779	16	Brazil	0.211	47	0.440	40	0.399	45
Australia	0.653	16	0.713	21	0.777	17	Argentina	0.141	54	0.374	45	0.397	46
United States	0.747	6	0.742	20	0.761	18	Colombia	0.217	46	0.348	48	0.368	47
United Arab Emirates	0.555	23	0.657	26	0.748	19	Thailand	0.181	49	0.316	52	0.358	48
Estonia	0.540	24	0.674	24	0.744	20	Kazakhstan	0.248	42	0.309	54	0.352	49
Canada	0.675	14	0.783	16	0.728	21	Mexico	0.174	50	0.384	44	0.331	50
Belgium	0.562	21	0.879	11	0.724	22	Peru	0.145	53	0.360	47	0.325	51
France	0.618	19	0.838	13	0.706	23	Ukraine	0.184	48	0.316	53	0.317	52
Portugal	0.560	22	0.761	19	0.693	24	Venezuela	0.036	57	0.290	56	0.317	53
Qatar	0.444	28	0.623	29	0.685	25	Philippines	0.156	52	0.316	51	0.309	54
Slovenia	0.433	29	0.628	28	0.679	26	Indonesia	0.129	55	0.283	57	0.297	55
Ireland	0.520	25	0.768	17	0.658	27	South Africa	0.117	56	0.348	49	0.270	56
Spain	0.477	26	0.674	25	0.658	28	India	0.163	51	0.290	55	0.260	57
Lithuania	0.420	30	0.614	30	0.644	29							

6.4 Generalising iSCI for Developing the UKCI

In the previous section, it has been illustrated that the FCM score has produced a more balanced and robust result than the LBG VQ and the PCA-FA/GME representations. In this section the FCM technique is generalised to produce the suggested *UKCI* baskets, sub-baskets and the final scores and ranks. This illustration can be viewed as a paradigm for aggregating multiple inputs to form a single and meaningful output as an index of indexes, hence, the *UKCI*. Even though the issue seems simple because of the nature that each of the selected indicators represents, however the methods for aggregating vast amounts of empirical data remain rather crude (Cherchye and Kuosmanen, 2004).

For the *iSCI* concept to be generalised to produce the *UKCI* final scores and sub-scores, and to answer the growing need for an 'all inclusive' unified indicator that encapsulates the major indices while keeping in mind that there are few shortcoming to that approach, therefore, a relation must be established between the indicators, to ensure that we are not aggregating heterogeneous variables.

6.4.1 Pre-Aggregation Tests and Analysis

To guarantee that the multi-dimensional inputs can be reduced to a lower dimensionality, without losing ground on preserving the differences in measurements between these indicators, it is suggested to examine the correlation and conduct PCA before zooming-in to produce the *UKCI* scores and sub-scores.

6.4.1.1 Correlation Analysis

A correlation analysis is conducted using the scores as reported by the six sources for 57 economies. The results presented in Table 6.7 reveal a high correlation between the IDI & KEI = (0.95); another strong correlation resulted between the NRI & GCI = (0.95), between the NRI & GII = (0.94) and between the GCI & WCY = (0.93). Also a high to moderate correlation between the WCY & NRI = (0.89), GII & IDI = (0.84) and (0.83) between WCY & GII has occurred. The lowest correlation is moderate which resulted between the WCY & KEI = (0.55). From this analysis it is confirmed that even though these indicators seem to differ

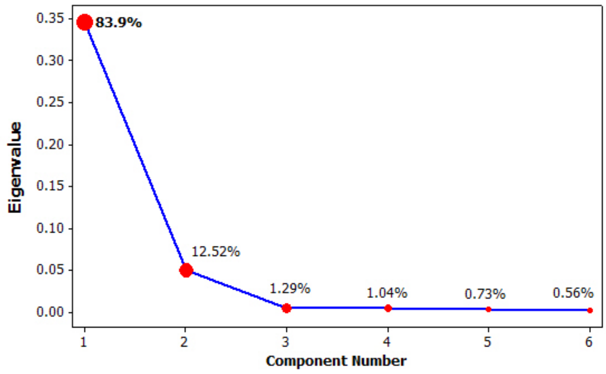


Figure 6.10: PCA eigenvalues and the number of principal components.

in names and purpose, they yield almost similar results in terms of knowledge performance and ranking between nations. Because of this redundancy, it should be possible to reduce the above indicators into a smaller number using any reduction methods like factor analysis, PCA or cluster analysis.

6.4.1.2 PCA Analysis

In response to the correlation results, the selected indicators were subjected to PCA analysis, to find patterns in data and to highlight their similarities and differences. Figure 6.10 shows the scree test result of the PCA analysis: the first component display the highest eigenvalue as it explains 83.86% of the variability in the data, and the second component accounts for 12.52%, so, the results of the scree test suggest that only the first two components are meaningful. Therefore, only the first two components were retained and a rotation will not be needed.

Table 6.7: Correlation coefficient matrix final scores, year 2011.

N= 57	IDI	KEI	GCI	GII	NRI	WCY
IDI	1.0					
KEI	0.95	1.0				
GCI	0.68	0.66	1.0			
GII	0.84	0.83	0.90	1.0		
NRI	0.79	0.75	0.95	0.94	1.0	
WCY	0.56	0.55	0.93	0.83	0.89	1.0

Combined, components 1 and 2 account for 96.38% of the variability in the data, which we can retain. At the bottom of the cliff lies the scree: the eigenvalues that represent the trivial components which adds to 3.62%, these can be comfortably discarded. Table 6.8 lists the complete result of the PCA analysis. These results confirms the very high similarity between these indices and the knowledge economy indicators tap slightly different variations in the underlying constructs. This milestone result justifies the original purpose for unification and confirms that we can aggregate these knowledge and competitiveness indices into a single value, without fearing that we are combining "apple and oranges". These results should allow us to smoothly carry out the early suggested fuzzy clustering model to produce the new Unified Macro-Knowledge Competitiveness Index (*UKCI*), which will bring these different indices together.

6.4.1.3 Normalisation

The same procedure as in Section. 6.3.1.4 is employed where all values are normalised to a range of $[0-1]$ where 0 implies poor and all the way to 1 as the top score.

6.4.1.4 ANFIS Predictions and Validation

To this end this research is proposing a unified index using fuzzy c-means clustering, which can provide us with the chance to model a complex issue like the rate of knowledge progress in a nation not based on any "personal expert judgement" or "skewed" statistical weighting, rather based on observed data from a well-

Table 6.8: Principal component analysis.

Eigenvalues of Covariance Matrix				
No.	Eigenvalue	% Total	Cumulative	Cumulative %
1	**5.03**	**83.86**	**5.03**	**83.86**
2	0.75	12.52	5.78	96.38
3	0.08	1.29	5.86	97.67
4	0.06	1.04	5.92	98.72
5	0.04	0.73	5.97	99.44
6	0.03	0.56	6.00	100.00

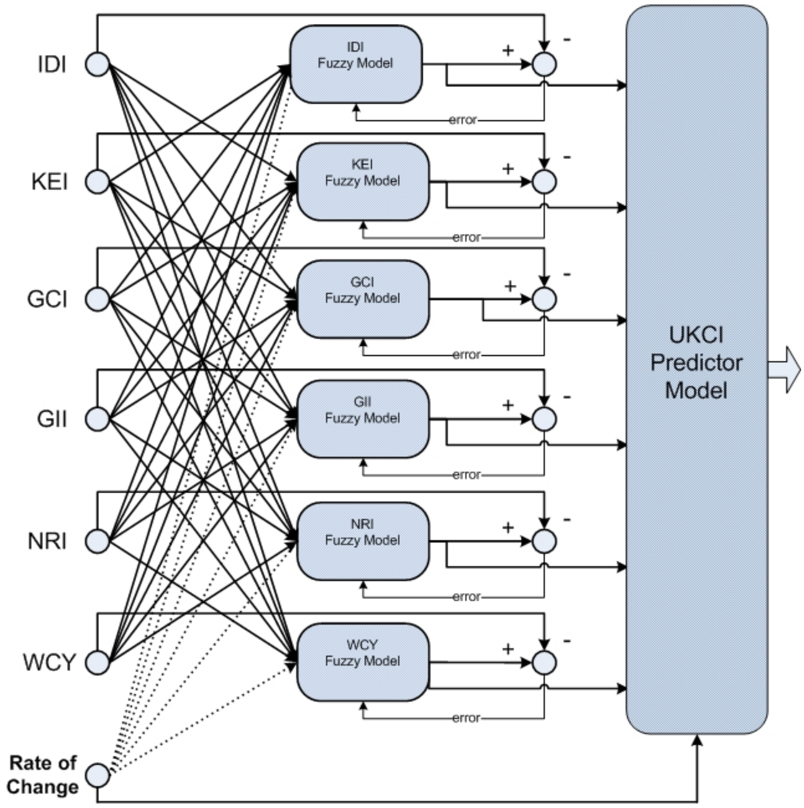


Figure 6.11: UKCI prediction model.

established indicators which can be fed into an Adaptive Neuro-Fuzzy Inference System (ANFIS) model to be trained and tested for overall validation.

To implement ANFIS, learning fuzzy models are developed initially. The schematic diagram shown in Figure 6.11 represents the ANFIS based macro-knowledge predictive model. This model will serve dual functions: As a KBE individual score predictor and a final score validator for the *UKCI* aggregator model. These six sub-models are designed to predict each index based on the available data from the other indices, (computing a model of a data set based on other data sets, making use of a composite evaluation function). Sub-models are constructed using multi-input single-output fuzzy rule based models. The last node in the model calculates the rate of change and stores the results. This node registers the final score X of country C at time t with respect to each of the presented six indicators I. Hence, three states can be identified which can be used as sub-inputs and as direct result of the partial derivatives: Sub-input1 $= X_t$, to identify Neutral Progress(NP), Sub-input2 = X_t - X_{t-1}, to identify Moving Progress (MP), and Sub-input3 = X_t - $2X_{t-1}$ + X_{t-2} to identify Accelerating Progress (AP).

6.4.1.5 Data Division for ANFIS Training and Testing

The dataset used in this section is collected from all the identified sources for creating the qualitative taxonomy as described in Chapter 5. The data consist of the final composite scores given originally by the six used composite indicators: the ITU-IDI, WB-KEI, WEF-GCI, INS-GII, WEF-NRI and IMD-WCY for 51 economies for three years periods 2009, 2010 and 2011. These final scores values were used to predict each of the used indicators in a multi-fold or permutation order.

To train the presented model, the dataset were separated into two parts: training set (51 economies final scores for 2009 and 2010) and testing set (51 economies final scores for 2011). To allow ANFIS to learn all probable states, so the inference system could generate high predictability rules input, it is suggested to shuffle the data in a random order so the datasets would have a good mix. ANFIS training is an iterative process, which calculates and minimises the sum of the squared differences between predictions and training instances. For the analysis, MATLAB fuzzy toolbox is employed. Using the collected data as input/output data set, Fuzzy Inference Systems (FIS) are constructed where membership function parameters are automatically tuned using either a backpropagation algorithm alone or in combination with a least squares type of method. This Hybrid adjustment allows the proposed fuzzy system to learn from the data and propose the rules to guide the proposed model and hence the desired aggregated and predicted output. Many trials were carried to achieve the most accurate prediction results. Parameters of the Knowledge Based Prediction (KBP) model are setup as follows: For the membership function, the Gaussian-bell shaped were used. To train the FIS, the hybrid learning algorithm were used and the sub clustering partition were utilised in order to generate the FIS method. Given separate sets of input and output data, the sub clustering function were modified to generate the FIS using FCM clustering. The function achieves this by attaining a set of

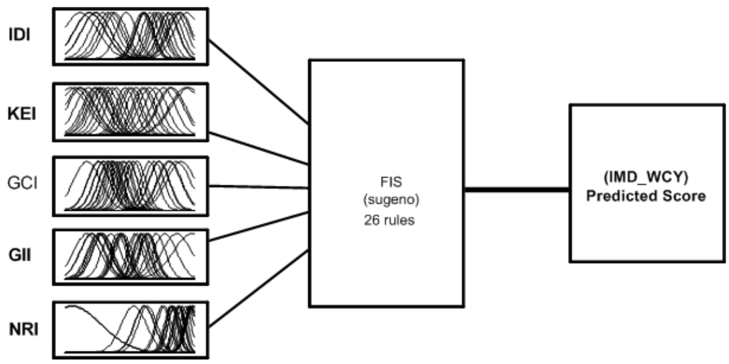


Figure 6.12: The inner-structure of IMD-WCY fuzzy sub-model.

rules that models the data behaviour. The rule attaining method first uses the FCM function to establish the number of rules and membership functions for the antecedents and consequents.

Figure 6.12, shows the details for the IMD-WCY fuzzy sub-model to illustrate the inputs data behaviour and the inner structure for one of these generated fuzzy inference systems, which is capable of predicting the IMD-WCY scores from the other indices using the ANFIS generated rules.

Sample of these rules which was created to give the decision for the IMD-WCY predicted scores are listed below.

Rule 1. If (IDI is cluster1) and (KEI is cluster1) and (GCI is cluster1) and (GII is cluster1) and (NRI is cluster1) then (IMD-WCY is cluster1)

Rule 2. If (IDI is cluster2) and (KEI is cluster2) and (GCI is cluster2) and (GII is cluster2) and (NRI is cluster2) then (IMD-WCY is cluster2)

Rule 3. If (IDI is cluster3) and (KEI is cluster3) and (GCI is cluster3) and (GII is cluster3) and (NRI is cluster3) then (IMD-WCY is cluster3)

.. ..

Rule 26. If (IDI is cluster26) and (KEI is cluster26) and (GCI is cluster26) and (GII is cluster26) and (NRI is cluster26) then (IMD-WCY is cluster26).

The full IMD-WCY fuzzy rule sub-model is presented in Figure 6.13 which depicts how the above rules are applied in order to generate a certain predicted IMD-WCY score.

The new scalable Unified Knowledge Progress Indicator is constructed to com-

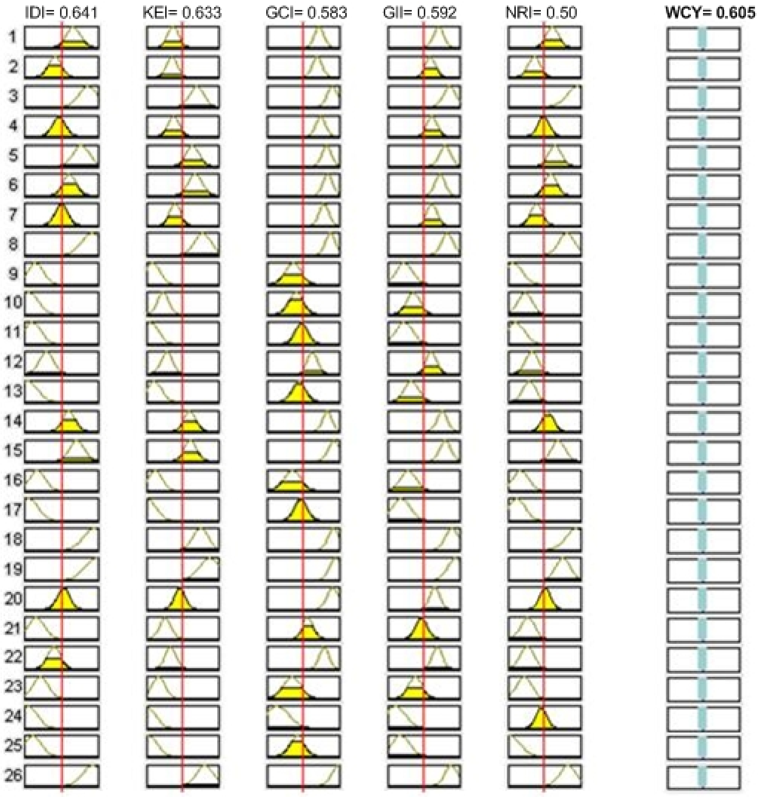


Figure 6.13: Fuzzy rules to construct the predicted IMD-WCY scores.

bine the knowledge and competitiveness indices into a new single meaningful index using the fuzzy clustering model which reflects the rate of knowledge competitiveness and progress in a nation; each sub-model represents a non-linear expression presented in a fuzzy rule-based format of the following form:

$$I_n = f\left(\sum I_i\right), i = [1, 2, 3, 4, 5, 6], i \neq n \tag{6.4}$$

where I_n is the predicted value for an index based on other indices I_i $i = [1, 2, 3, 4, 5, 6], i \neq n$. These Sub-models are trained to minimise the total error. Root Mean Square (RMS) error is used to measure the error of prediction for each index. The goal is to minimise the total error as indicated below:

$$RMSE = \sqrt{\frac{\sum e^2_{IDI} + e^2_{KEI} + e^2_{GCI} + e^2_{GII} + e^2_{NRI} + e^2_{WCY}}{6}} \tag{6.5}$$

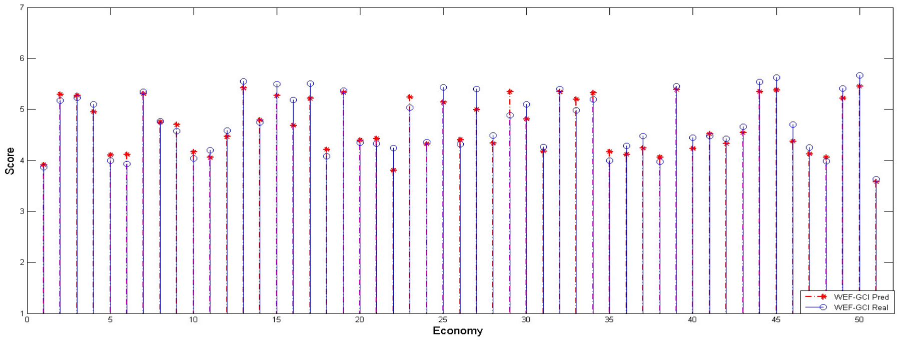
Table 6.9: UKCI overall predicted errors.

Predicted Knowledge Indicators		
Input	**Predicted**	**RMSE**
IDI, KEI, GII, NRI, WCY	GCI	0.168
KEI, GCI, GII, NRI, WCY	IDI	0.191
IDI, GCI, GII, NRI, WCY	KEI	0.201
IDI, KEI, GCI, NRI, WCY	GII	0.397
IDI, KEI, GCI, GII, WCY	NRI	0.368
IDI, KEI, GCI, NRI, GII	WCY	0.311
Predicted UKCI Root Mean Sq. Error		
IDI, KEI, GCI, GII, NRI, WCY	**UKCI**	**0.2725**

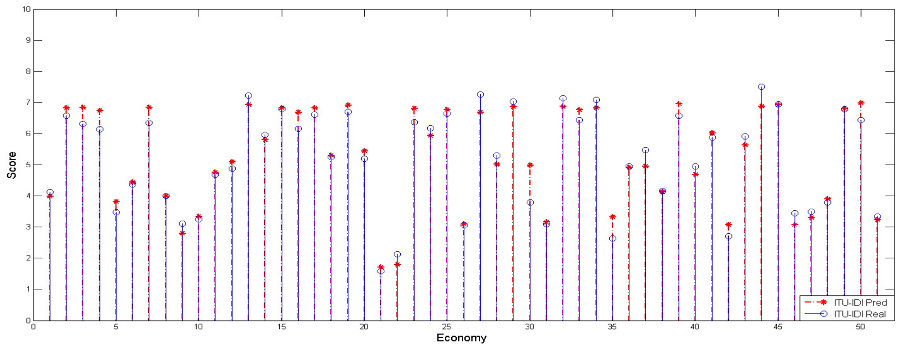
where e_{IDI}, e_{KEI}, e_{GCI}, e_{GII}, e_{NRI}, and e_{WCY} are the average prediction error for each fuzzy sub-model. Using ANFIS for the six sub-models, where sub-clustering technique is used to generate the initial rules.

6.4.1.6 ANFIS Predictions and Validation Results

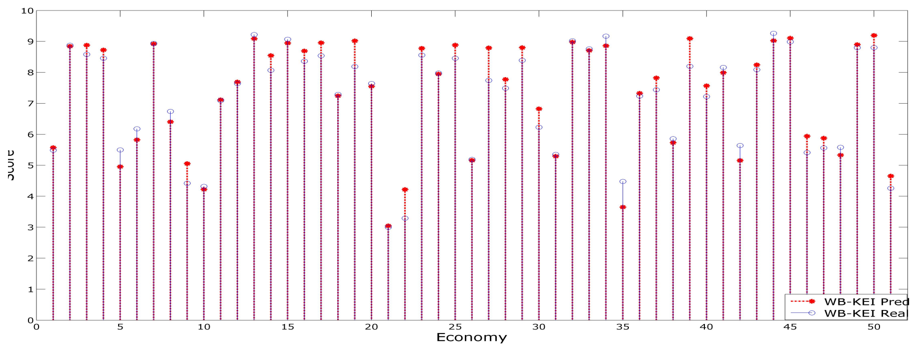
The aggregated error is first summed up for the final output using the RMSE measure. Table 6.9 summarizes the average errors obtained by the model and the overall error is calculated using expression 6.5. Therefore by aggregating the indices the new model would predict an aggregated value which will form the proposed *UKCI* with a margin of combined error = 0.2725. The overall fit is good for all indices, but the best fit is achieved for the WEF-GCI score with an average error of 0.168 as shown in Figure 6.14(a). The second best result was achieved for predicting the ITU-IDI score as Figure 6.14(b) shows. The third best result was achieved for the WB-KEI as Figure 6.14(c) shows. The worst predicted value by the model is presented in Figure Figure 6.15(a) for the INS-GII with an average error of 0.397. This is due to the nature of formation as it depends on "soft variables" to form its final score and its innovation focused. Figure 6.15(b) and (c) show the plot for the predicted scores for the WEF-NRI and IMD-WCY respectively.



(a) GCI

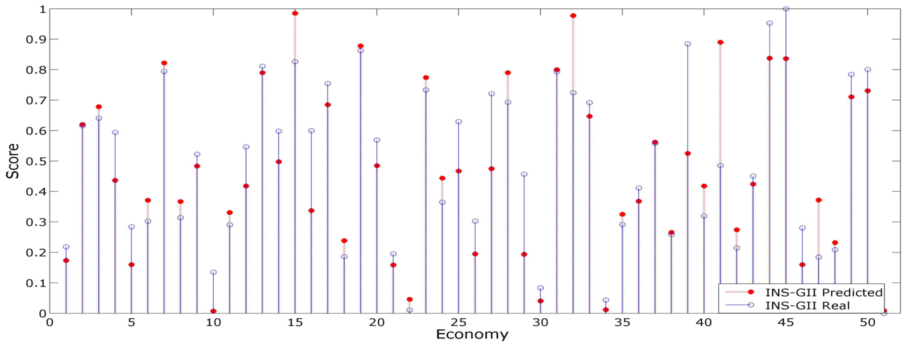


(b) IDI

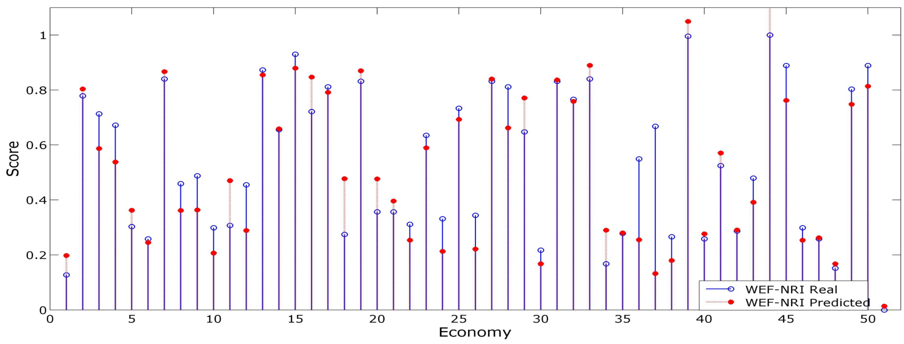


(c) KEI

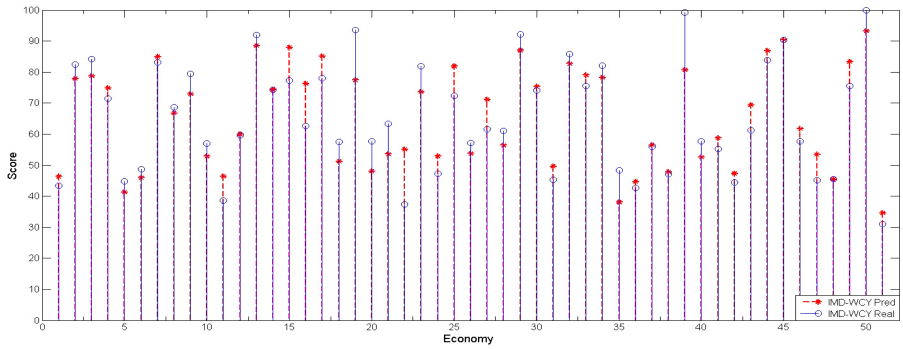
Figure 6.14: Predicted vs. 2011 real scores. (a) WEF-GCI , (b) ITU-IDI, (c) WB-KEI.



(a) GII



(b) NRI



(c) WCY

Figure 6.15: Predicted vs. 2011 real scores. (a) INS-GII, (b) WEF-NRI, (c) IMD-WCY.

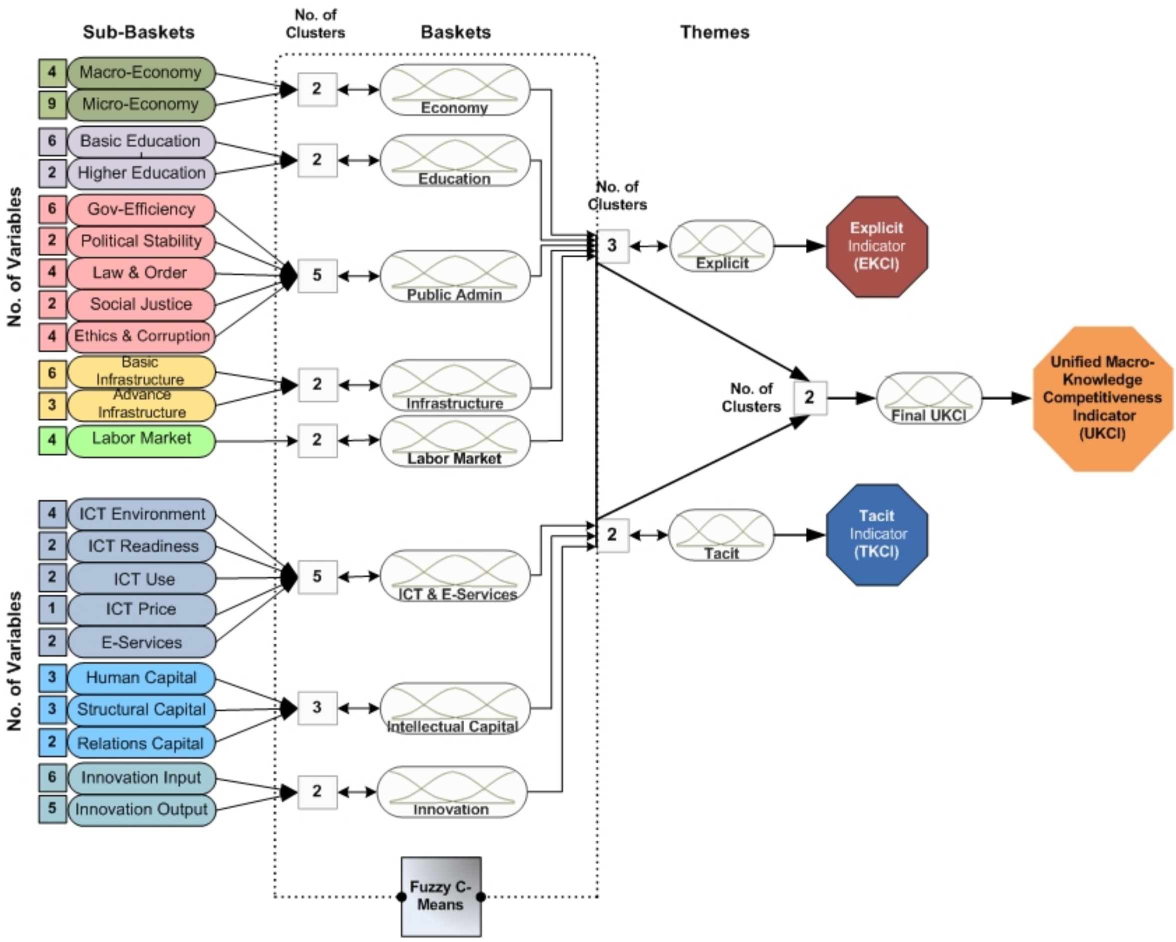


Figure 6.16: Schematic diagram of the UKCI fuzzy inference model.

6.4.2 UKCI Fuzzy Inference Aggregation Model

As a last step and to form the *UKCI*, it is proposed to cluster the available data for each economy and to identify a cluster centre to represent the unified index for that economy, using the FCM procedure as described in Section 5.7.1 Equations 5.12 to 5.16. The FCM clustering space as well as the distance function used with the FCM algorithm can be used in this case and as follows: Each score point will have a degree of membership to the scores clusters, as in fuzzy logic, rather than belonging entirely to just one cluster. Therefore, scores on the border of a cluster, could be in the cluster to a smaller degree than points in the middle of cluster. Cluster centres are determined first at the learning stage, and then the

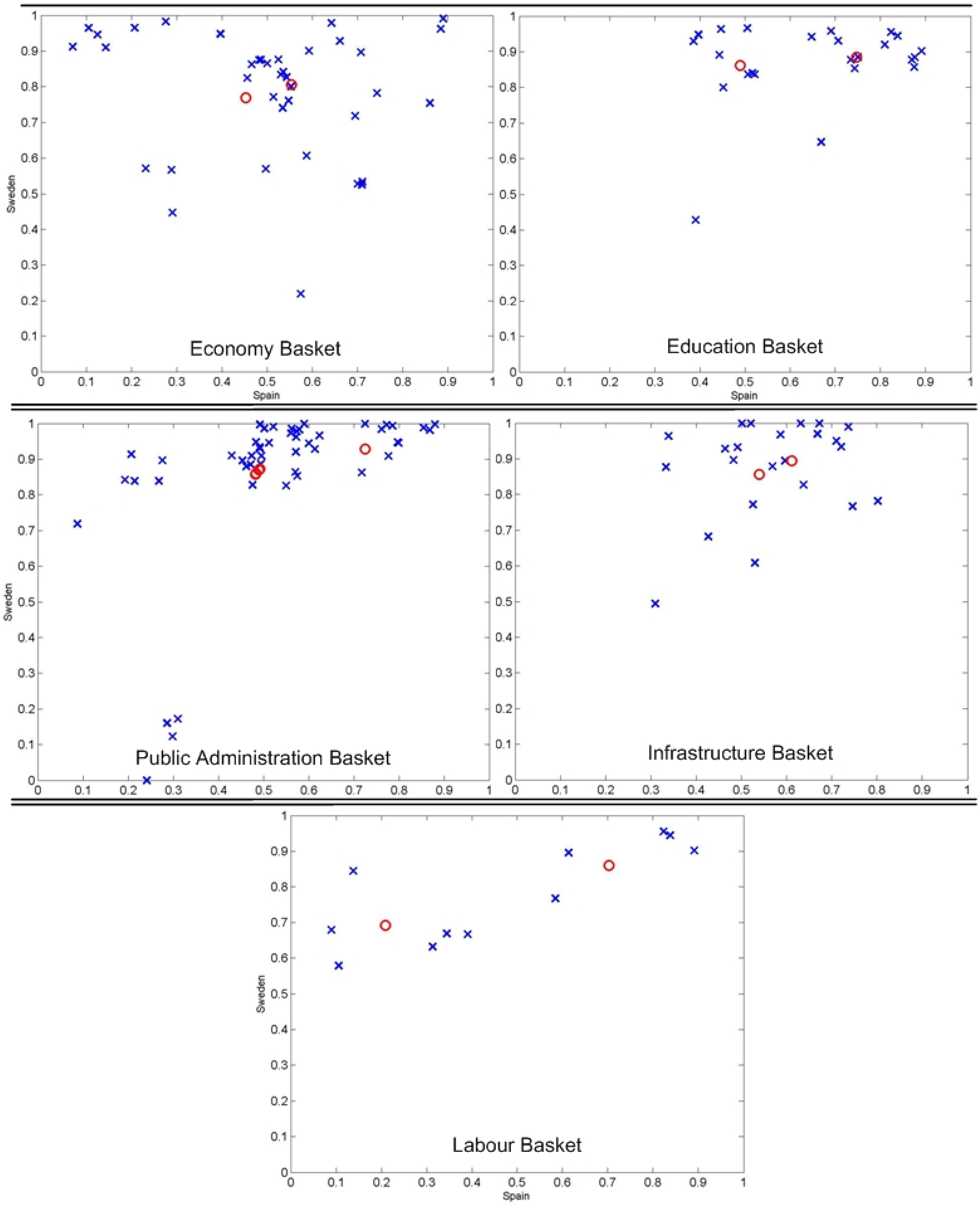


Figure 6.17: FCM aggregation process of the five baskets scores to one explicit theme score - Spain vs. Sweden, year 2011-2012.

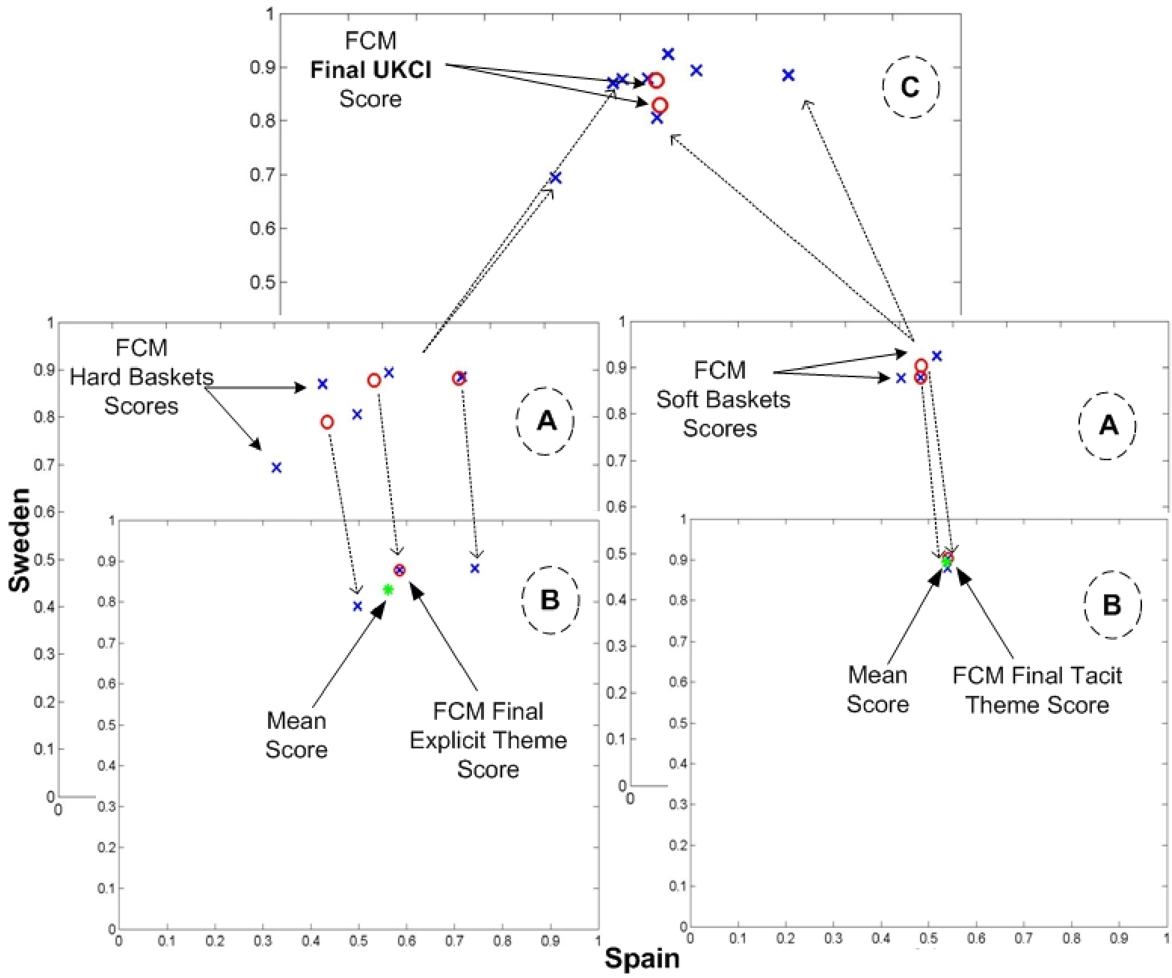


Figure 6.18: FCM aggregation process for UKCI tacit, explicit and final scores - Spain vs. Sweden, year 2011-2012.

classification is made by the comparison of distance between the incoming features and each cluster centre. Figure 6.16 shows the suggested macro-knowledge index aggregation model using FCM clustering to generate the *UKCI* scores and ranks for both the explicit, tacit and the final *UKCI*. Considering the limitation of the dataset (data records) and rather large features (economies), the number of clusters is dependant on the number of input features. These clusters are representing a range where the predicted value belongs. Figures 6.17 and 6.19 respectively depict the process of FCM aggregation to obtain the *UKCI* explicit, tacit and final scores for two randomly selected economies, happen to be (Spain vs.

Sweden for the year 2011-2012). This sample illustrates the aggregation process for optimum value (the cluster centroid marked as a red circle) for the sampled economies, where a fuzzy membership is given to highlight the importance of each index.

6.5 Assessing the UKCI Applicability for the MENA Region

To assess the *UKCI* applicability, effectiveness and added values, this section is focused on generating the *UKCI* scores and ranks for the Middle East and North Africa (MENA) region. The MENA region is selected for logistic, strategic and practical reasons. The analysis on this region is intended to highlight the sophistication of analysis that can be achieved by using the *UKCI* detailed ranks and scores for countries with different cultural, socio-economic, and technical conditions.

Despite the facts that most of the region economies enjoys massive amount of wealth represented in natural gas and oil, its member states are still considered underdeveloped or developing. A close examination of 15 of the region states, as listed in Table 6.10, shows the amount of missing data points, out of the *UKCI* 80 source variables. This is a generic problem facing many underdeveloped and developing nations. This problem is a major issue for countries like Iraq which has 92% of its data points missing. For reliability and control the threshold for missing points has been set for 50% for three consecutive years. Therefore Iraq and many other economies have been dropped out of the *UKCI* aggregation. However, under this criteria it was possible to impute missing points for Libya which has 79% of its data points missing for one year, but it meets the threshhold condition for the three years period. The FCM Optimal Completion Strategy (OCS) as described in Section 5.8, is utilised to impute the missing data points and FCM algorithm to weight and aggregate the variables for the MENA region in addition to the other 57 economies which enjoy complete datasets. Table 6.11 shows the imputed data points after applying the OCS technique.

Table 6.10: MENA region normalised row variables, year 2011-2012.

Baskets	Sub-Baskets	No.	Indicators	Algeria	Bahrain	Egypt	Iraq	Jordan	Kuwait	Lebanon	Libya	Morocco	Oman	Qatar	Saudi Arabia	Tunisia	UAE	Yemen
ECONOMY		1	GCI_MacroEco	0.764	0.627	0.287	NaN	0.453	0.973	0.323	NaN	0.747	0.947	0.928	0.853	0.670	0.865	0.308
		2	GCI_Goods Mrkt	0.183	0.877	0.321	NaN	0.537	0.541	0.638	NaN	0.470	0.716	0.802	0.881	0.571	0.851	0.216
	Macro	3	KEI_Economic Regime Pillar	0.232	0.689	0.459	NaN	0.580	0.602	0.436	NaN	0.476	0.717	0.708	0.583	0.387	0.669	0.292
		4	WCY_Economic Performance	NaN	NaN	NaN	NaN	0.153	NaN	NaN	NaN	NaN	NaN	0.938	NaN	NaN	0.478	NaN
		5	GCI_Financial Mrkt	0.116	0.801	0.431	NaN	0.525	0.539	0.541	NaN	0.536	0.674	0.757	0.785	0.489	0.660	0.000
		6	GCI_Mrkt size	0.551	0.286	0.625	NaN	0.346	0.462	0.417	NaN	0.496	0.414	0.466	0.651	0.457	0.564	0.396
	Micro	7	NRI_Market Environment	0.137	0.806	0.465	NaN	0.488	0.555	0.545	0.080	0.478	0.666	0.803	0.739	0.518	0.749	NaN
		8	WCY_Productivity & Efficiency	NaN	NaN	NaN	NaN	0.000	NaN	NaN	NaN	NaN	NaN	1.000	NaN	NaN	0.052	NaN
		9	WCY_Finance	NaN	NaN	NaN	NaN	0.310	NaN	NaN	NaN	NaN	NaN	0.569	NaN	NaN	0.293	NaN
		10	WCY_Management Practices	NaN	NaN	NaN	NaN	0.121	NaN	NaN	NaN	NaN	NaN	0.655	NaN	NaN	0.224	NaN
		11	WCY_Attitudes and Values	NaN	NaN	NaN	NaN	0.345	NaN	NaN	NaN	NaN	NaN	0.828	NaN	NaN	0.431	NaN
		12	GII_Business Environment	0.572	0.958	0.814	NaN	0.701	0.836	0.612	NaN	0.761	0.904	0.929	0.955	0.704	0.914	0.487
		13	GII_Market Sophistication	0.250	0.568	0.273	NaN	0.408	0.439	0.329	NaN	0.264	0.274	0.331	0.520	0.263	0.516	0.147
EDUCATION		14	GCI_Health & Primary Education	0.674	0.848	0.638	NaN	0.718	0.700	0.835	NaN	0.643	0.680	0.876	0.747	0.827	0.819	0.326
		15	NRI_Individual readiness	0.623	0.843	0.629	NaN	0.745	0.716	0.757	0.481	0.339	0.728	0.875	0.748	0.835	0.896	NaN
	Basic	16	KEI_Education Pillar	0.529	0.686	0.332	NaN	0.558	0.366	0.554	NaN	0.197	0.525	0.336	0.568	0.454	0.584	0.150
		17	IDI_ICT Skills	0.578	0.772	0.525	NaN	0.714	NaN	0.714	NaN	0.331	0.636	0.579	0.679	0.651	0.679	0.299
		18	WCY_Education	NaN	NaN	NaN	NaN	0.224	NaN	NaN	NaN	NaN	NaN	0.328	NaN	NaN	0.259	NaN
		19	GII_Education	0.655	0.823	0.573	NaN	0.750	0.718	0.587	NaN	0.621	0.592	0.581	0.834	0.738	0.652	0.552
	Higher	20	GCI_Higher Education	0.383	0.739	0.366	NaN	0.579	0.459	0.651	NaN	0.409	0.557	0.648	0.694	0.660	0.701	0.093
		21	GII_Tertiary education	0.277	0.671	0.102	NaN	0.398	0.313	0.477	NaN	0.413	0.366	0.372	0.336	0.234	0.494	0.087
PUBLIC ADMINISTRATION		22	GCI_Institutions	0.187	0.778	0.369	NaN	0.531	0.523	0.228	NaN	0.423	0.789	0.805	0.827	0.575	0.756	0.043
		23	NRI_Political & Regulatory Environt	0.142	0.575	0.377	NaN	0.497	0.320	0.119	0.000	0.414	0.510	0.599	0.643	0.533	0.578	NaN
	Gov-Efficincy	24	KEI_Gov't Effectivness	0.374	0.632	0.403	0.226	0.336	0.522	0.423	0.229	0.461	0.630	0.598	0.481	0.543	0.552	0.269
		25	WCY_Institiotional Framework	NaN	NaN	NaN	NaN	0.379	NaN	NaN	NaN	NaN	NaN	0.672	NaN	NaN	0.569	NaN
		26	WCY_Fiscal Policy	NaN	NaN	NaN	NaN	0.759	NaN	NaN	NaN	NaN	NaN	0.983	NaN	NaN	1.000	NaN
		27	WCY_Public Finance	NaN	NaN	NaN	NaN	0.362	NaN	NaN	NaN	NaN	NaN	0.793	NaN	NaN	0.931	NaN
	Political	28	KEI_Political Stability	0.391	0.585	0.464	0.174	0.432	0.747	0.332	0.644	0.547	0.803	0.876	0.608	0.678	0.795	0.185
	Stability	29	GII_Political Environment	0.286	0.496	0.379	NaN	0.375	0.641	0.360	NaN	0.409	0.672	0.709	0.366	0.447	0.711	0.031
		30	KEI_Rule of Law	0.379	0.654	0.527	0.184	0.515	0.676	0.401	0.330	0.510	0.704	0.695	0.589	0.577	0.562	0.313
	Law & Order	31	KEI_Regulatory Quality	0.298	0.740	0.520	0.316	0.524	0.602	0.571	0.297	0.536	0.679	0.609	0.596	0.559	0.564	
		32	GII_Regulatory Environment	0.270	0.746	0.543	NaN	0.634	0.720	0.469	NaN	0.410	0.751	0.785	0.644	0.538	0.737	0.318
		33	WCY_Business Legislation	NaN	NaN	NaN	NaN	0.414	NaN	NaN	NaN	NaN	NaN	0.655	NaN	NaN	0.603	NaN
	Social Justice	34	KEI_Voice & Accountability	0.313	0.354	0.262	0.302	0.250	0.445	0.491	0.077	0.376	0.302	0.209	0.115	0.226	0.233	0.242
		35	WCY_Societal Framework	NaN	NaN	NaN	NaN	0.276	NaN	NaN	NaN	NaN	NaN	0.241	NaN	NaN	0.397	NaN
	Ethics	36	GCI_Ethics & Corruptions	0.187	0.778	0.369	NaN	0.531	0.523	0.228	NaN	0.423	0.789	0.805	0.827	0.575	0.756	0.043
	&	37	KEI_Control of Corruption	0.305	0.483	0.287	0.100	0.354	0.506	0.218	0.116	0.383	0.512	0.765	0.458	0.390	0.614	0.144
	Corruption	38	WCY_Bribing & Corruption	NaN	NaN	NaN	NaN	0.424	NaN	NaN	NaN	NaN	NaN	0.814	NaN	NaN	0.612	NaN
		39	AMN_Corruption Preciption Index	0.170	0.449	0.164	0.030	0.343	0.386	0.117	0.057	0.238	0.414	0.695	0.357	0.278	0.651	0.065
INFRASTRUCTURE		40	GCI_Infrastructure	0.356	0.680	0.430	NaN	0.493	0.556	0.196	NaN	0.458	0.695	0.697	0.725	0.538	0.855	0.126
		41	NRI_Infrastructure environment	0.222	0.473	0.316	NaN	0.354	0.439	0.354	0.260	0.280	0.334	0.529	0.424	0.404	0.610	NaN
		42	IDI_ICT Access	0.302	0.716	0.391	NaN	0.421	NaN	0.369	NaN	0.388	0.504	0.759	0.672	0.333	0.719	0.129
	Basic	43	WCY_Basic Infrastructure	NaN	NaN	NaN	NaN	0.000	NaN	NaN	NaN	NaN	NaN	0.517	NaN	NaN	0.621	NaN
		44	WCY_Health & Environment	NaN	NaN	NaN	NaN	0.224	NaN	NaN	NaN	NaN	NaN	0.483	NaN	NaN	0.362	NaN
		45	GII_General Infrastructure	0.285	0.754	0.199	NaN	0.435	0.408	0.376	NaN	0.204	0.499	0.416	0.510	0.230	0.710	0.172
		46	WCY_Technological Infrastructure	NaN	NaN	NaN	NaN	0.172	NaN	NaN	NaN	NaN	NaN	0.345	NaN	NaN	0.517	NaN
	Advance	47	WCY_Scientific Infrastructure	NaN	NaN	NaN	NaN	0.172	NaN	NaN	NaN	NaN	NaN	0.138	NaN	NaN	0.103	NaN
		48	GII_Energy	0.335	0.365	0.268	NaN	0.478	0.444	0.251	NaN	0.183	0.145	0.349	0.484	0.202	0.293	0.233
LABOUR		49	GCI_Labor Mrkt Efficincy	0.173	0.648	0.101	NaN	0.355	0.508	0.352	NaN	0.208	0.570	0.645	0.550	0.355	0.622	0.231
	Labour	50	KEI_Edu & Human Resorce	0.529	0.686	0.332	NaN	0.558	0.366	0.554	NaN	0.197	0.525	0.336	0.568	0.454	0.584	0.150
		51	WCY_Labor Market	NaN	NaN	NaN	NaN	0.172	NaN	NaN	NaN	NaN	NaN	0.828	NaN	NaN	0.293	NaN
		52	GII_Knowledge Workers	0.211	0.359	0.376	NaN	0.198	0.189	0.619	NaN	0.153	0.298	0.442	0.410	0.113	0.726	0.268
ICT & E-SERVICES		53	NRI_Infrastructure environment	0.222	0.473	0.316	NaN	0.354	0.439	0.354	0.260	0.280	0.334	0.529	0.424	0.404	0.610	NaN
	ICT Envirmt	54	IDI_ICT Access	0.302	0.716	0.391	NaN	0.421	NaN	0.369	NaN	0.388	0.504	0.759	0.672	0.333	0.719	0.129
		55	WCY_Technological Infrastructure	NaN	NaN	NaN	NaN	0.172	NaN	NaN	NaN	NaN	NaN	0.345	NaN	NaN	0.517	NaN
		56	GII_ICT Infrastructure	0.125	0.523	0.299	NaN	0.261	0.385	0.326	NaN	0.356	0.211	0.212	0.469	0.316	0.499	0.053
		57	GCI_Technological Readiness	0.162	0.560	0.278	NaN	0.399	0.370	0.297	NaN	0.370	0.464	0.623	0.524	0.401	0.657	0.060
	Readiness	58	NRI_Gov't Readiness	0.274	0.740	0.469	NaN	0.577	0.311	0.000	0.306	0.449	0.743	0.971	0.746	0.814	0.883	NaN
		59	NRI_Individual Usage	0.241	0.683	0.327	NaN	0.407	0.440	0.297	0.229	0.366	0.540	0.685	0.609	0.409	0.748	NaN
	Use	60	IDI_ICT Use	0.069	0.409	0.151	NaN	0.192	NaN	0.162	NaN	0.259	0.323	0.476	0.456	0.192	0.651	0.046
	Price	61	IDI_ICT Price	0.962	0.994	0.955	NaN	0.950	NaN	0.962	NaN	0.870	0.987	0.966	0.989	0.969	0.999	0.488
		62	NRI_Gov't Usage	0.132	0.787	0.510	NaN	0.535	0.342	0.184	0.238	0.295	0.469	0.515	0.424	0.585	0.508	NaN
	E-Services	63	KEI_ICT	0.403	1.000	0.304	NaN	0.458	0.674	0.354	NaN	0.401	0.669	0.769	0.873	0.496	0.503	0.092
INTELLECTUAL CAPITAL		64	GCI_Business Sophistication	0.146	0.599	0.401	NaN	0.418	0.458	0.501	NaN	0.390	0.539	0.817	0.771	0.499	0.713	0.160
	Human	65	NRI_Business Readiness	0.374	0.417	0.248	NaN	0.228	0.149	0.543	0.000	0.305	0.490	0.715	0.566	0.570	0.685	NaN
		66	GII_Knowledge Workers	0.211	0.359	0.376	NaN	0.198	0.189	0.619	NaN	0.153	0.298	1.000	0.410	0.113	0.726	0.268
		67	GII_Knowledge Creation	0.024	0.031	0.083	NaN	0.274	0.058	0.183	NaN	0.067	0.036	0.010	0.021	0.136	0.017	0.007
	Structural	68	GII_Creative Intangibles	0.127	0.438	0.534	NaN	0.945	0.570	0.505	NaN	0.405	0.664	1.000	0.925	0.970	0.967	0.171
		69	GII_Research & Development	0.101	0.288	0.122	NaN	0.195	0.119	0.220	NaN	0.161	0.492	0.715	0.168	0.318	0.539	NaN
		70	GCI_Business Sophistication	0.146	0.599	0.401	NaN	0.418	0.458	0.501	NaN	0.390	0.539	0.817	0.771	0.499	0.713	0.160
	Relations	71	WCY_Labour Market	NaN	NaN	NaN	NaN	0.172	NaN	NaN	NaN	NaN	NaN	0.828	NaN	NaN	0.293	NaN
INNOVATION		72	GCI_Innovation	0.169	0.372	0.284	NaN	0.342	0.323	0.244	NaN	0.328	0.430	0.736	0.606	0.465	0.557	0.000
		73	NRI_Business Usage	0.000	0.327	0.233	NaN	0.267	0.201	0.214	0.135	0.217	0.330	0.428	0.399	0.418	0.437	NaN
	Innov Input	74	KEI_Innovation	0.273	0.396	0.338	NaN	0.331	0.466	0.425	NaN	0.288	0.542	0.604	0.342	0.437	0.625	0.091
		75	WCY_Scientific Infrastructure	NaN	NaN	NaN	NaN	0.172	NaN	NaN	NaN	NaN	NaN	0.138	NaN	NaN	0.103	NaN
		76	GII_Innovation Linkages	0.088	0.379	0.217	NaN	0.449	0.249	0.236	NaN	0.142	0.917	0.740	0.604	0.416	0.910	NaN
		77	GII_Knowledge Absorption	0.214	0.116	0.152	NaN	0.253	0.039	0.417	NaN	0.199	0.296	NaN	0.261	0.159	0.000	0.083
		78	GII_Knowledge Impact	0.159	0.553	0.237	NaN	0.211	0.343	NaN	NaN	0.256	0.351	1.000	0.352	0.237	0.357	0.334
	Innov Output	79	GII_Knowledge Diffusion	0.000	0.225	0.253	NaN	0.279	0.912	0.571	NaN	0.328	0.349	NaN	0.213	0.286	0.004	0.246
		80	GII_Creative Goods & Services	0.000	0.043	0.250	NaN	0.459	0.015	0.483	NaN	0.142	0.036	0.005	0.040	0.126	0.359	0.010

Table 6.11: MENA region imputed variables using FCM-OCS, year 2011-2012.

Baskets	Sub-Baskets	No.	Indicators	Algeria	Bahrain	Egypt	Iraq	Jordan	Kuwait	Lebanon	Libya	Morocco	Oman	Qatar	Saudi Arabia	Tunisia	UAE	Yemen
ECONOMY		1	GCI_MacroEco	0.7665	0.61352	0.23508	NaN	0.42027	1	0.27534	0.22817	0.74771	0.97048	0.94901	0.8658	0.66183	0.87922	0.25924
		2	GCI_Goods Mrkt	0.182836	0.876866	0.320896	NaN	0.537313	0.541045	0.63806	0.22817	0.470149	0.716418	0.802239	0.880597	0.570896	0.850746	0.216418
	Macro	3	KEI_Economic Regime Pillar	0.20985	0.67984	0.44377	NaN	0.56773	0.59037	0.42005	0.22817	0.46101	0.70895	0.69925	0.57097	0.36939	0.65936	0.27237
		4	WCY_Economic Performance	0.259701	0.544292	0.34476	NaN	0.153219	0.476807	0.398419	0.22817	0.335108	0.499168	0.938485	0.514191	0.453972	0.477998	0.158011
		5	GCI_Financial Mrkt	0.116022	0.801105	0.430939	NaN	0.524862	0.538674	0.541436	0.22817	0.535912	0.674033	0.756906	0.78453	0.48895	0.660221	0
		6	GCI_Mrkt size	0.47935	0.17142	0.56444	NaN	0.2403	0.37603	0.32336	0.22817	0.41452	0.31931	0.38008	0.59483	0.36995	0.49353	0.29905
	Micro	7	NRI_Market Environment	0.137124	0.80602	0.464883	NaN	0.488294	0.555184	0.545151	0.080268	0.478261	0.665552	0.802676	0.73913	0.518395	0.749164	0.158011
		8	WCY_Productivity & Efficiency	0.259701	0.544292	0.34476	NaN	0	0.476808	0.398419	0.22817	0.335108	0.499168	1	0.514191	0.453972	0.051724	0.158011
		9	WCY_Finance	0.259701	0.544292	0.34476	NaN	0.310345	0.476808	0.398419	0.22817	0.335108	0.499168	0.568966	0.514191	0.453972	0.293103	0.158011
		10	WCY_Management Practices	0.259701	0.544292	0.34476	NaN	0.12069	0.476808	0.398419	0.22817	0.335108	0.499168	0.655172	0.514191	0.453972	0.224138	0.158011
		11	WCY_Attitudes and Values	0.259701	0.544292	0.34476	NaN	0.344828	0.476808	0.398419	0.22817	0.335108	0.499168	0.827586	0.514191	0.453972	0.431034	0.158011
		12	GII_Business Environment	0.572238	0.957507	0.814448	NaN	0.701133	0.835694	0.611898	0.22817	0.760623	0.903683	0.929178	0.954674	0.703966	0.913598	0.487252
		13	GII_Market Sophistication	0.25035	0.567832	0.272727	NaN	0.408392	0.439161	0.328671	0.22817	0.264336	0.274126	0.331469	0.52028	0.262937	0.516084	0.146853
EDUCATION		14	GCI_Health & Primary Education	0.55	0.78929	0.5	NaN	0.61071	0.58571	0.77143	0.22817	0.50714	0.55714	0.82857	0.65	0.76071	0.75	0.06786
		15	NRI_Individual readiness	0.623188	0.843478	0.628986	NaN	0.744928	0.715942	0.756522	0.481159	0.33913	0.727536	0.875362	0.747826	0.834783	0.895652	0.158011
	Basic	16	KEI_Education Pillar	0.5011	0.66703	0.29231	NaN	0.53187	0.32857	0.52747	0.22817	0.14945	0.4967	0.2967	0.54286	0.42198	0.55934	0.1
		17	IDI_ICT Skills	0.49145	0.72507	0.42877	NaN	0.65527	0.476809	0.65527	0.22817	0.19516	0.56125	0.49288	0.61396	0.57977	0.61396	0.1567
		18	WCY_Education	0.259701	0.544292	0.344761	NaN	0.224138	0.476808	0.398419	0.22817	0.335108	0.499168	0.327586	0.514191	0.453972	0.258621	0.158011
		19	GII_Education	0.64771	0.81913	0.56435	NaN	0.74522	0.71219	0.5785	0.22817	0.61311	0.58322	0.57221	0.83014	0.73263	0.64456	0.54233
	Higher	20	GCI_Higher Education	0.36383	0.73123	0.34657	NaN	0.56603	0.44274	0.64	0.228171	0.39096	0.54383	0.63753	0.68438	0.64986	0.69178	0.06547
		21	GII_Tertiary education	0.276619	0.670692	0.102086	NaN	0.398463	0.312843	0.477497	0.22817	0.412733	0.365532	0.372119	0.335895	0.233809	0.493963	0.086718
PUBLIC		22	GCI_Institutions	0.186992	0.777778	0.368564	NaN	0.531165	0.523035	0.227642	0.22817	0.422764	0.788618	0.804878	0.826558	0.574526	0.756098	0.04336
ADMINISTRATION		23	NRI_Political & Regulatory Environt	0.141643	0.575071	0.376771	NaN	0.497006	0.320113	0.11898	0	0.413598	0.509915	0.598802	0.643059	0.532578	0.577844	0.158011
	Gov-Efficincy	24	KEI_Gov't Effectivness	0.374087	0.631764	0.40326	0.226	0.33605	0.521662	0.422542	0.229232	0.461442	0.629675	0.598363	0.481213	0.542582	0.55173	0.269178
		25	WCY_Institiotional Framework	0.259701	0.544292	0.34476	NaN	0.37931	0.476807	0.398419	0.22817	0.335108	0.499168	0.672414	0.514191	0.453972	0.568966	0.158011
		26	WCY_Fiscal Policy	0.259701	0.544292	0.344761	NaN	0.758621	0.476808	0.398419	0.22817	0.335108	0.499168	0.982759	0.514191	0.453972	1	0.158011
		27	WCY_Public Finance	0.259701	0.544292	0.344761	NaN	0.362069	0.476808	0.398419	0.22817	0.335108	0.499168	0.793103	0.514191	0.453972	0.931034	0.158011
	Political	28	KEI_Political Stability	0.391035	0.584504	0.464045	0.174	0.431719	0.747132	0.332188	0.644467	0.546759	0.802552	0.875729	0.607985	0.677515	0.794671	0.18477
	Stability	29	GII_Political Environment	0.286334	0.495662	0.378525	NaN	0.375174	0.640998	0.360087	0.22817	0.408894	0.672451	0.708508	0.36551	0.446855	0.711297	0.031453
		30	KEI_Rule of Law	0.379251	0.653957	0.527037	0.184	0.515311	0.675527	0.401124	0.329937	0.509694	0.703984	0.694757	0.588641	0.577422	0.561519	0.312699
	Law & Order	31	KEI_Regulatory Quality	0.297807	0.739806	0.520491	0.316	0.524077	0.6025	0.570843	0.297358	0.535926	0.678677	0.609474	0.595971	0.558635	0.564214	0.158011
		32	GII_Regulatory Environment	0.269637	0.745604	0.54279	NaN	0.634232	0.719812	0.468933	0.22817	0.410317	0.751465	0.785463	0.643611	0.538101	0.737397	0.317702
		33	WCY_Business Legislation	0.259701	0.544292	0.34476	NaN	0.413793	0.476808	0.398419	0.22817	0.335108	0.499168	0.655172	0.514191	0.453972	0.603448	0.158011
	Social Justice	34	KEI_Voice & Accountability	0.313384	0.353621	0.262269	0.302	0.24997	0.445204	0.490696	0.077366	0.3763	0.302254	0.20918	0.114888	0.225579	0.232888	0.242054
		35	WCY_Societal Framework	0.259701	0.544292	0.34476	NaN	0.275862	0.476808	0.398419	0.22817	0.335108	0.499168	0.241379	0.514191	0.453972	0.396552	0.158011
	Ethics	36	GCI_Ethics & Corruptions	0.186992	0.777778	0.368564	NaN	0.531165	0.523035	0.227642	0.22817	0.422764	0.788618	0.804878	0.826558	0.574526	0.756098	0.04336
	&	37	KEI_Control of Corruption	0.304642	0.482798	0.286633	0.100	0.35443	0.506344	0.218022	0.115926	0.383481	0.512017	0.764552	0.4575	0.390109	0.614266	0.143862
	Corruption	38	WCY_Bribing & Corruption	0.259701	0.544292	0.34476	NaN	0.424156	0.476808	0.398419	0.22817	0.335108	0.499168	0.813674	0.514191	0.453972	0.611733	0.158011
		39	AMN_Corruption Preciption Index	0.169722	0.448937	0.164381	0.030	0.342822	0.386416	0.116862	0.057062	0.237963	0.414084	0.695318	0.357377	0.277954	0.651429	0.0654
INFRASTRUCTURE		40	GCI_Infrastructure	0.26292	0.63371	0.34831	NaN	0.42022	0.49213	0.0809	0.22817	0.37978	0.65169	0.65393	0.68539	0.47191	0.83371	0
		41	NRI_Infrastructure environment	0.20753	0.46331	0.30345	NaN	0.34227	0.42906	0.34227	0.24636	0.26691	0.32172	0.52041	0.41307	0.39252	0.60262	0.158011
		42	IDI_ICT Access	0.23223	0.68725	0.33021	NaN	0.36377	0.476809	0.30605	0.22817	0.32753	0.45504	0.73557	0.63893	0.26713	0.69128	0.04297
	Basic	43	WCY_Basic Infrastructure	0.259701	0.544292	0.34476	NaN	0	0.476808	0.398419	0.22817	0.335108	0.499168	0.517241	0.514191	0.453972	0.62069	0.158011
		44	WCY_Health & Environment	0.259701	0.544292	0.34476	NaN	0.224138	0.476808	0.398419	0.22817	0.335108	0.499168	0.482759	0.514191	0.453972	0.362069	0.158011
		45	GII_General Infrastructure	0.285486	0.754386	0.199362	NaN	0.435407	0.408293	0.376396	0.22817	0.204147	0.499203	0.416268	0.510367	0.229665	0.709729	0.172249
		46	WCY_Technological Infrastructure	0.259701	0.544292	0.34476	NaN	0.172414	0.476808	0.398419	0.22817	0.335108	0.499168	0.344828	0.514191	0.453972	0.517241	0.158011
	Advance	47	WCY_Scientific Infrastructure	0.259701	0.544292	0.34476	NaN	0.172414	0.476808	0.398419	0.22817	0.335108	0.499168	0.137931	0.514191	0.453972	0.103448	0.158011
		48	GII_Energy	0.33452	0.365362	0.26809	NaN	0.478055	0.443654	0.251483	0.22817	0.182681	0.144721	0.348754	0.483986	0.201661	0.293001	0.232503
LABOUR		49	GCI_Labor Mrkt Efficincy	0.172638	0.648208	0.100977	NaN	0.355049	0.508143	0.351792	0.22817	0.208469	0.570033	0.644951	0.550489	0.355049	0.62215	0.23127
	Labor	50	KEI_Edu & Human Resorce	0.5011	0.66703	0.29231	NaN	0.53187	0.32857	0.52747	0.22817	0.14945	0.4967	0.2967	0.54286	0.42198	0.55934	0.1
		51	WCY_Labor Market	0.259701	0.544292	0.34476	NaN	0.172414	0.476808	0.398419	0.22817	0.335108	0.499168	0.827586	0.514191	0.453972	0.293103	0.158011
		52	GII_Knowledge Workers	0.210956	0.358974	0.376457	NaN	0.198135	0.188811	0.618881	0.22817	0.152681	0.298368	0.441725	0.410256	0.113054	0.726107	0.268065
ICT & E-SERVICES		53	NRI_Infrastructure environment	0.20753	0.46331	0.30345	NaN	0.34227	0.42906	0.34227	0.24636	0.26691	0.32172	0.52041	0.41307	0.39252	0.60262	0.158011
	ICT Envirmt	54	IDI_ICT Access	0.23223	0.68725	0.33021	NaN	0.36377	0.476809	0.30605	0.22817	0.32753	0.45504	0.73557	0.63893	0.26713	0.69128	0.04297
		55	WCY_Technological Infrastructure	0.259701	0.544292	0.34476	NaN	0.172414	0.476808	0.398419	0.22817	0.335108	0.499168	0.344828	0.514191	0.453972	0.517241	0.158011
		56	GII_ICT Infrastructure	0.11607	0.51786	0.29241	NaN	0.25335	0.37835	0.3192	0.22817	0.34933	0.20313	0.20424	0.46317	0.30915	0.49442	0.04353
		57	GCI_Technological Readiness	0.161836	0.560386	0.277778	NaN	0.398551	0.369565	0.297101	0.22817	0.369565	0.463768	0.623188	0.524155	0.400966	0.657005	0.060386
	Readiness	58	NRI_Gov't Readiness	0.274286	0.74	0.468571	NaN	0.577143	0.311429	0	0.305714	0.448571	0.742857	0.971429	0.745714	0.814286	0.882857	0.158011
		59	NRI_Individual Usage	0.241309	0.683027	0.327198	NaN	0.406953	0.439673	0.296524	0.229039	0.366053	0.539877	0.685072	0.609407	0.408998	0.748466	0.158011
	Use	60	IDI_ICT Use	0.068966	0.408685	0.150702	NaN	0.191571	0.476808	0.162197	0.22817	0.259259	0.323116	0.476373	0.455939	0.191571	0.651341	0.045977
	Price	61	IDI_ICT Price	0.93235	0.98998	0.91982	NaN	0.9098	0.476809	0.93235	0.22817	0.76698	0.97745	0.93987	0.97996	0.94488	0.99749	0.08547
		62	NRI_Gov't Usage	0.131519	0.786848	0.510204	NaN	0.535147	0.342404	0.183673	0.238095	0.294785	0.469388	0.514739	0.424036	0.585034	0.507937	0.158011
	E-Services	63	KEI_ICT	0.403471	1	0.303688	NaN	0.457701	0.673536	0.353579	0.22817	0.401302	0.669197	0.76898	0.873102	0.495662	0.503254	0.092191
INTELLECTUAL		64	GCI_Business Sophistication	0.13427	0.59328	0.39283	NaN	0.41026	0.45093	0.49451	0.22817	0.38121	0.53227	0.81407	0.76759	0.4916	0.70949	0.1488
CAPITAL	Human	65	NRI_Business Readiness	0.374172	0.417219	0.248344	NaN	0.228477	0.149007	0.543046	0	0.304636	0.490066	0.715232	0.566225	0.569536	0.68543	0.158011
		66	GII_Knowledge Workers	0.210956	0.358974	0.376457	NaN	0.198135	0.188811	0.618881	0.22817	0.152681	0.298368	1	0.410256	0.113054	0.726107	0.268065
		67	GII_Knowledge Creation	0.023632	0.031095	0.083333	NaN	0.273632	0.058458	0.182836	0.22817	0.067164	0.03607	0.00995	0.021144	0.135572	0.017413	0
	Structural	68	GII_Creative Intangibles	0.127159	0.437991	0.533752	NaN	0.945055	0.569859	0.505495	0.22817	0.405024	0.66405	1	0.924647	0.970173	0.967033	0.171115
		69	GII_Research & Development	0.08261	0.27335	0.10418	NaN	0.17798	0.10078	0.2041	0.22817	0.14392	0.48113	0.70934	0.15073	0.30401	0.52995	0.158011
		70	GCI_Business Sophistication	0.13427	0.59328	0.39283	NaN	0.41026	0.45093	0.49451	0.22817	0.38121	0.53227	0.81407	0.76759	0.4916	0.70949	0.1488
	Relations	71	WCY_Labor Market	0.259701	0.544292	0.34476	NaN	0.172414	0.476808	0.398419	0.22817	0.335108	0.499168	0.827586	0.514191	0.453972	0.293103	0.158011
INNOVATION		72	GCI_Innovation	0.168704	0.371638	0.283619	NaN	0.342298	0.322738	0.244499	0.22817	0.327628	0.430318	0.735941	0.606357	0.464548	0.557457	0
		73	NRI_Business Usage	0	0.327044	0.232704	NaN	0.267296	0.201258	0.213836	0.13522	0.216981	0.330189	0.427673	0.399371	0.418239	0.437107	0.158011
	Innov Input	74	KEI_Innovation	0.22738	0.35819	0.29707	NaN	0.28973	0.43276	0.38875	0.22817	0.24328	0.51345	0.57946	0.30073	0.4022	0.60147	0.03423
		75	WCY_Scientific Infrastructure	0.259701	0.544292	0.34476	NaN	0.172414	0.476808	0.398419	0.22817	0.335108	0.499168	0.137931	0.514191	0.453972	0.103448	0.158011
		76	GII_Innovation Linkages	0.07477	0.37009	0.20561	NaN	0.44112	0.23738	0.2243	0.22817	0.12897	0.91589	0.73645	0.59813	0.40748	0.90841	0.158011
		77	GII_Knowledge Absorption	0.21371	0.115591	0.151882	NaN	0.252688	0.038978	0.416667	0.22817	0.198925	0.295699	0.604268	0.260753	0.158602	0	0.083333
		78	GII_Knowledge Impact	0.03125	0.4851	0.1211	NaN	0.09115	0.2432	0.398419	0.22817	0.14299	0.25242	1	0.25357	0.1211	0.25933	0.23284
	Innov Output	79	GII_Knowledge Diffusion	0	0.225061	0.253041	NaN	0.278589	0.912409	0.57056	0.22817	0.328467	0.349148	0.604268	0.212895	0.285888	0.00365	0.245742
		80	GII_Creative Goods & Services	0	0.043046	0.25	NaN	0.458609	0.014901	0.483444	0.171492	0.142384	0.036424	0.004967	0.039735	0.125828	0.359272	0.009934

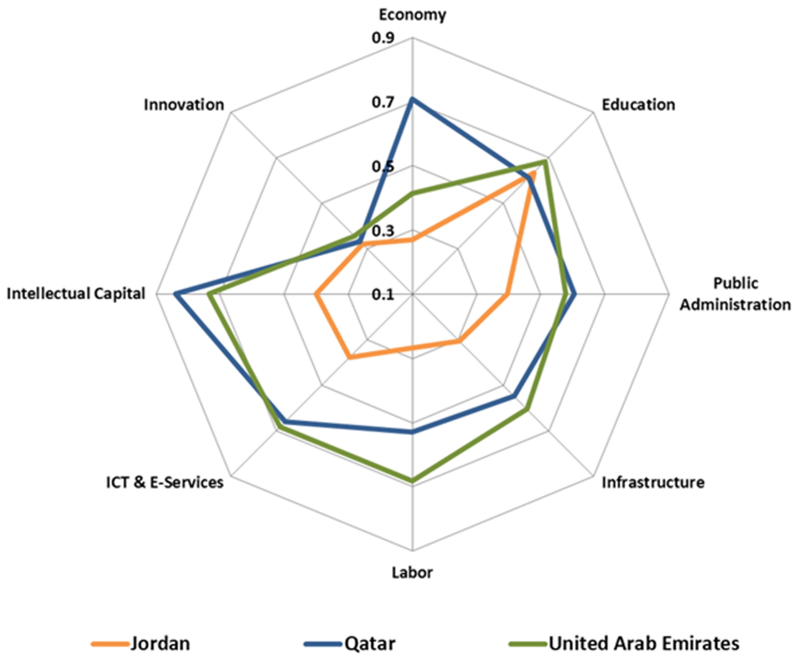


Figure 6.19: UKCI baskets scores comparison, Jordan vs. Qatar vs. UAE, year 2011-2012.

6.5.1 The UKCI Drill Down Capability for Comprehensive Decision Making

By tapping into the achieved scores and ranks for the MENA region from the presented tables in Appendix B, many illustrative ways can be introduced to better visualise the results. For example the results from the *UKCI* index detailed baskets can be charted for couple of countries for comparisons and detail analysis, as depicted by the radar chart for three countries (Jordan, Qatar and United Arab Emirates) in Figure 6.19. This scores comparisons reveal that all three MENA countries have a couple of progress issues that should be regarded when devising policies. For example when it comes to Qatar, the biggest prospective for policy improvements should focus on the area of education. Strategies need to consider: primary and higher education (instituting an improved curriculum, building more schools, building the ranks and reputation of their national universities and research centres etc.). While these facts might have been revealed by many other frameworks, the *UKCI* provides a unified and all-inclusive results that associates

Table 6.12: UKCI final scores - MENA region, years 2009, 2010 and 2011.

Country/Economy	UKCI_Final 2009-2010 Score	Rank
Qatar	0.581	27
United Arab Emirates	0.577	28
Kuwait	0.496	37
Saudi Arabia	0.490	38
Bahrain	0.471	41
Tunisia	0.458	43
Oman	0.445	45
Jordan	0.424	49
Turkey	0.397	51
Lebanon	0.384	54
Iran	0.328	62
Egypt	0.312	64
Morocco	0.308	65
Syria	0.253	68
Algeria	0.224	69
Libya	0.188	71
Yemen	0.130	73

Country/Economy	UKCI_Final 2010-2011 Score	Rank	Status
Qatar	0.602	26	1+
United Arab Emirates	0.595	27	1+
Saudi Arabia	0.503	33	5+
Bahrain	0.487	35	6+
Oman	0.460	39	6+
Kuwait	0.440	43	6-
Tunisia	0.423	44	1-
Turkey	0.365	50	1+
Lebanon	0.352	53	1+
Jordan	0.349	54	5-
Egypt	0.337	57	7+
Iran	0.314	61	1+
Morocco	0.311	62	3+
Algeria	0.242	68	1+
Syria	0.221	69	1-
Libya	0.204	70	1+
Yemen	0.156	73	*

Country/Economy	UKCI_Final 2011-2012 Score	Rank	Status
Qatar	0.585	28	2-
United Arab Emirates	0.577	30	3-
Bahrain	0.541	33	2+
Saudi Arabia	0.533	35	2-
Oman	0.479	42	3-
Kuwait	0.439	44	1-
Turkey	0.407	49	1+
Tunisia	0.407	50	6-
Lebanon	0.400	51	2+
Jordan	0.359	57	3-
Egypt	0.337	62	5-
Morocco	0.327	65	3-
Iran	0.326	66	5-
Syria	0.268	68	1+
Algeria	0.251	69	1-
Libya	0.198	71	1-
Yemen	0.151	73	*

such policies to KBE development and competitiveness.

Furthermore, the nature of the *UKCI* framework allows policymakers to arrange policy proposals via the allocated learned weighted structure employed. This ensures a natural, robust and data-driven analysis based on each country or collated profile in order to develop different policies at variable degrees. For example, although the three countries need more effective strategies relating to innovation, their priority levels, as measured by their respective standardised scores, vary. Therefore, increasing the innovation level becomes a more immediate strategy priority for both Jordan and the United Arab Emirates than Qatar. Furthermore, by examining the temporal progress from Table 6.12 and by following each of these countries scores and ranks over time, Jordan has lost a total of 8 ranks from 2009 to 2011 with a constant decline in almost every aspect except education. Figure 6.20 presents the aggregated tacit and explicit knowledge competitiveness indicators TKCI and EKCI for Jordan from year 2009 to 2011. The details of the tacit score are expanded in a pie chart to highlight each basket contributions towards the making of the final tacit aggregated score for the year 2011.

Such and more detailed analysis can be further developed using the achieved detailed scores and ranks to draw a wider picture and transparent policies to benefit local citizens, decision makers and foreign investors.

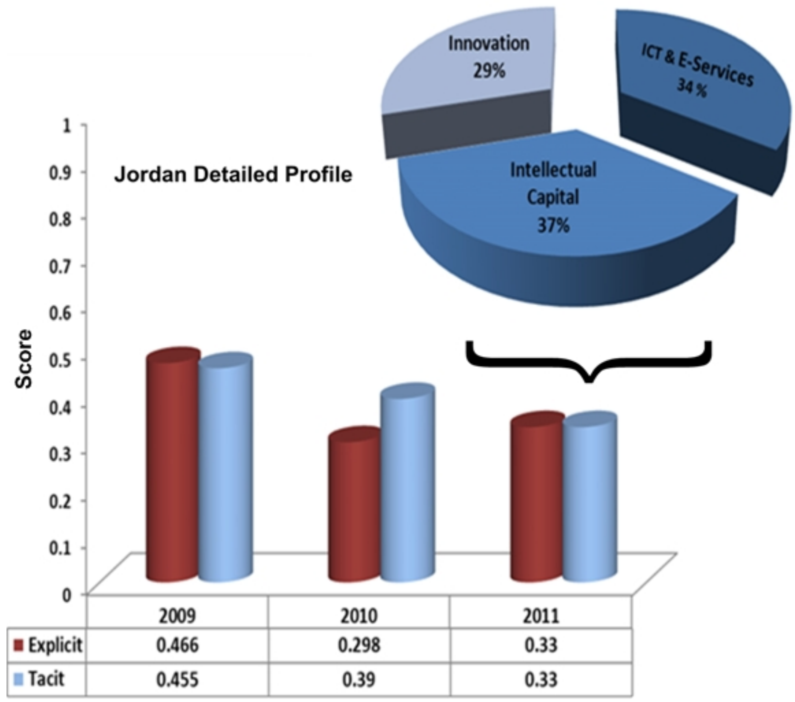


Figure 6.20: Jordan detailed tacit and explicit profile, years 2009-2011.

6.6 UKCI Themes and Final Measurements

Table 6.13 lists the *UKCI*, EKCI and the TKCI final scores and ranks for 73 economies for the year 2011-2012. The full results for the previous years 2009 and 2010, in addition to the *UKCI* 8 baskets and 2 themes detailed results for the three available years are listed in Appendix B, Tables. 2, 3, 4, 5, 6, 7, 8 and 9. The presented results for the *UKCI* themes and baskets enables the present research to explain many trends within and between nations for effective KBE developments and progress.

Table 6.13: UKCI final aggregated scores for 73 countries, year 2011-2012.

Countries	UKCI _Final 2011-2012		Explicit Knowledge 2011-2012		Tacit Knowledge 2011-2012	
	Score	Rank	Score	Rank	Score	Rank
Sweden	**0.881**	1	0.872	4	0.898	1
Switzerland	**0.878**	2	0.911	1	0.869	2
Singapore	**0.850**	3	0.884	2	0.821	3
Denmark	**0.845**	4	0.878	3	0.762	9
Finland	**0.837**	5	0.853	5	0.811	5
Hong Kong	**0.810**	6	0.816	8	0.732	14
Netherlands	**0.810**	7	0.818	7	0.807	7
Canada	**0.798**	8	0.826	6	0.751	13
United States	**0.797**	9	0.784	12	0.808	6
Germany	**0.785**	10	0.774	15	0.799	8
Norway	**0.776**	11	0.796	11	0.688	18
Luxembourg	**0.771**	12	0.805	10	0.705	16
United Kingdom	**0.768**	13	0.783	13	0.751	12
Taiwan China	**0.755**	14	0.685	22	0.817	4
Australia	**0.746**	15	0.776	14	0.587	24
Iceland	**0.741**	16	0.816	9	0.592	23
Austria	**0.738**	17	0.763	16	0.673	20
Belgium	**0.734**	18	0.744	18	0.670	21
Japan	**0.734**	19	0.732	19	0.753	10
Israel	**0.711**	20	0.680	23	0.751	11
Ireland	**0.711**	21	0.705	21	0.726	15
New Zealand	**0.706**	22	0.751	17	0.540	29
Korea Rep.	**0.700**	23	0.649	24	0.704	17
France	**0.695**	24	0.721	20	0.677	19
Estonia	**0.612**	25	0.623	25	0.540	28
Malaysia	**0.606**	26	0.582	29	0.592	22
Czech Republic	**0.599**	27	0.609	27	0.569	27
Qatar	**0.585**	28	0.563	31	0.389	40
Spain	**0.578**	29	0.584	28	0.514	31
United Arab Emirates	**0.577**	30	0.613	26	0.399	39
Portugal	**0.550**	31	0.558	33	0.482	34
Slovenia	**0.550**	32	0.558	32	0.494	32
Bahrain	**0.541**	33	0.564	30	0.291	62
Italy	**0.533**	34	0.494	39	0.526	30
Saudi Arabia	**0.533**	35	0.499	36	0.367	51
Chile	**0.530**	36	0.550	34	0.381	46
Hungary	**0.515**	37	0.494	38	0.581	25
China	**0.504**	38	0.443	42	0.574	26
Lithuania	**0.501**	39	0.510	35	0.385	43
Poland	**0.498**	40	0.483	40	0.435	37
Slovak Republic	**0.483**	41	0.494	37	0.375	48
Oman	**0.479**	42	0.475	41	0.382	45
Thailand	**0.440**	43	0.385	46	0.453	35
Kuwait	**0.439**	44	0.433	44	0.363	53
Croatia	**0.437**	45	0.436	43	0.373	50
Greece	**0.434**	46	0.431	45	0.374	49
Brazil	**0.431**	47	0.353	53	0.487	33
Russian Federation	**0.411**	48	0.374	47	0.406	38
Turkey	**0.407**	49	0.365	50	0.376	47
Tunisia	**0.407**	50	0.372	49	0.350	54
Lebanon	**0.400**	51	0.356	51	0.389	41
Bulgaria	**0.393**	52	0.373	48	0.320	59
Romania	**0.392**	53	0.339	55	0.384	44
South Africa	**0.380**	54	0.355	52	0.365	52
India	**0.366**	55	0.290	63	0.452	36
Mexico	**0.360**	56	0.314	59	0.346	56
Jordan	**0.359**	57	0.330	56	0.330	58
Kazakhstan	**0.359**	58	0.343	54	0.226	68
Colombia	**0.357**	59	0.302	60	0.333	57
Argentina	**0.356**	60	0.280	64	0.387	42
Ukraine	**0.343**	61	0.280	65	0.347	55
Egypt	**0.337**	62	0.317	57	0.270	64
Philippines	**0.329**	63	0.272	66	0.312	60
Peru	**0.329**	64	0.294	61	0.216	69
Morocco	**0.327**	65	0.315	58	0.250	66
Iran	**0.326**	66	0.292	62	0.279	63
Indonesia	**0.310**	67	0.236	69	0.302	61
Syria	**0.268**	68	0.267	67	0.180	70
Algeria	**0.251**	69	0.253	68	0.085	72
Venezuela	**0.237**	70	0.143	73	0.266	65
Libya	**0.198**	71	0.169	72	0.148	71
Mauritania	**0.190**	72	0.220	70	0.240	67
Yemen	**0.151**	73	0.174	71	0.085	73

6.7 Summary

A framework and an empirical case study for developing future Third Generation Intelligent Synthetic Composite Indicators is proposed. The *iSCI* framework and the empirical case study presented in this chapter is successfully constructed using computational intelligent methods to combine the efforts of eleven complex, multi-dimensional variables, into a new Unified ICT Index. After rigours validation and robustness analysis, the methods used in the case study were then generalised to build the Unified Macro-Knowledge Competitiveness Indicator (*UKCI*). A detailed analysis for a full region was also presented to highlight the importance of the proposed methods to impute missing values, weight and aggregate the variables which can be instantly mobilised as a robust methods to produce not only the *UKCI* but any future intelligent synthetic composite indicators for any other fields. In the next chapter this study is extended to investigate and devise a suitable forecasting framework of the near future behaviour of KBE competitiveness progress in a nation.

Chapter 7

Forecasting Macro-Knowledge Competitiveness

7.1 Introduction

In today's economy, nations are competing in many aspects, including innovation and knowledge progress. Even though there are many composite indicators to measure knowledge and innovation on a micro and macro levels, nonetheless, benefits to decision makers still limited due to numerous progress indicators, without any unified, easy to visualize and evaluate forecasting capabilities.

Measuring and predicting the future directions of KBE on a macro scale and competitive level has gained momentum. A similar concept is the primary concern of the "State of the Future Index" (Freudenberg, 2003), which is a quantitative time series, 10-year outlook, that indicates the changing state of the future and shows whether conditions promise to get better or worse (Gordon, 2005). However, a study by Chen et al. (2009) argues that knowledge is intangible and cannot be easily quantified or predicted, but its effect or outcome can be assessed. Furthermore, predicting the behaviour of a person or society is the most complicated example of prognosis and that the future developments of countries belong among the most complicated cases (Dostál, 1998).

Many contemporary studies presented computational intelligence techniques to deal with the prediction of complicated and uncertain cases, for example Ilonen et al. (2006) integrated Bass model, Artificial Neural Network(ANN), and Kohonen Self-Organizing Map (SOM) to forecast the diffusion of innovations in nations. A study by Sarle (1994) utilised a combination of computationally intelligent methods including multiple agent-based knowledge integration mechanism, fuzzy cognitive map, particle and swarm optimization to aggregate knowledge from multiple experts to solve a target problem.

In this chapter, a novel approach on integration of computational intelligence and traditional prediction techniques to forecast the future performance of KBE is presented. The focus is to forecast any of the selected knowledge indicators for any given economy in presence of limited data availability but with the required high accuracy. This is achieved by feeding ANN with panel data structure and also utilising unsupervised learning techniques to cluster the predicted results.

The remainder of this chapter is organised as follows. In Section 7.2, the KBE progress forecasting framework is presented which includes the proposed forecasting methodologies for the multiple regression, panel data analysis and the ANN based KBE forecasting models. In Section 7.3, the results for indicators prediction and mapping economies are presented. The conclusions are summarised in Section 7.4

7.2 Macro-Knowledge Competitiveness Forecasting Framework

When the progress and competitiveness of KBE in a specific economy is forecast, the quality of the framework process will depend on the availability and quality of the measurement tools, the reliability, and the data range of the macro-knowledge indicators. The proposed framework to present a one-step ahead forecast of six major KBE related indices is shown in Figure 7.1. The hierarchy of this KBE forecasting framework is carried out in the following order; at the highest level is the data set, which consists of the so far collected macro-knowledge synthetic aggregated indices for the forecasting models under consideration. The next step down is to test the degree and rate of change over time for each of the worked on indicators. To achieve that, it is suggested to correlate each indicator on itself

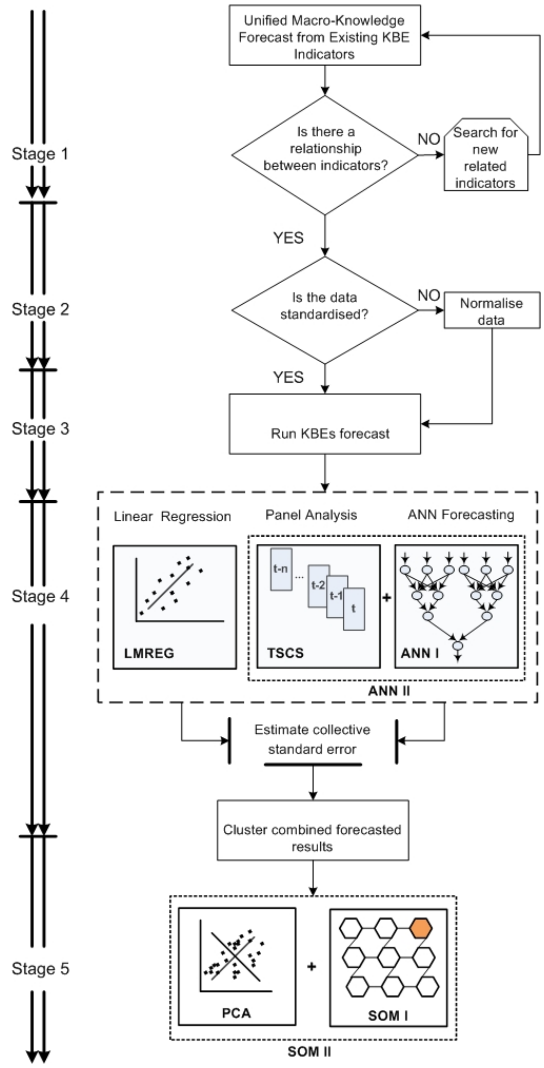


Figure 7.1: Proposed KBE progress forecasting framework.

over the available time frame. If there is a strong relationship or a sensitive one, it is recommended to experiment with couple of forecasting process to select the one that produce the most accurate results. The forecasting step is the core of this framework and in this step three major forecasting techniques will be tested; Linear Multiple Regression (LMREG), TSCS, and ANN are used to compare the performance results. This study is also suggesting to feed the ANN with panel data set to test whether this will have an impact on the ANN forecasting performance results. The standard error for all models will be estimated using the Mean Absolute Percent Error (MAPE) and the Mean Square Error (MSE). Finally, the winning performance results will be clustered using Kohonen's Self-Organizing Map, then visualised and evaluated. The framework exhibited in Figure 7.1 is developed as an aid to macro-knowledge indicators selection, to forecast and categories homogeneous KBE's for strategic decisions making and policy support. More details about each stage of this framework are explained below.

7.2.0.1 Panel Data Analysis

There are several models for the analysis of panel data or the time-series cross-sectional (TSCS) regressions, such as the constant coefficients model (pooled), fixed effect, random effect, dynamic panel, robust panel, covariance structure model etc. After careful examination and extensive consideration for each and every one of these models, the selections have been narrowed between three major models which could best suit the collected data set for this study, and therefore they would produce a meaningful analysis. The selected types are:

- Pooled model or the constant coefficients model - This considered to be the simplest model as it has a constant intercepts and slopes, usually there is neither large number of entities nor large temporal effects; When this is the case it is advised to run the early mentioned typical linear regression model.

- Fixed effects model (FEM) - The fixed effects model is the first and earliest model; it is also called the least square with dummy variables (LSDV) and can be represented as follows:

$$Y_{it} = \alpha_i + \beta' x_{it} + \cdots + \varepsilon_{it} \tag{7.1}$$

  where i = 1,2. . . N ; t = 1,2. . . T ; the slopes α_i are constant and are not associated with time like managerial skills, tacit knowledge, know-how etc., but the intercepts are different giving there is a heterogeneity between the selected countries.

- Random effects model (REM) - This model is similar, however, the slopes α_is are treated as random variables rather than fixed constants. This model is also known as the Variance Components Model (VCM).

The next section proceeds with different tests to illustrate how to examine for the presence of statistically related KBEs and/or time series effects, and to establish which is more appropriate for the case at hand, the random or the fixed effects model?

7.2.0.2 Fixed vs. Random Effects Models

Whether to treat the effects of α_i as fixed or random number's is not an easy question to answer. It does make a significant difference in the estimates of the parameters especially when only few observations are available for different economies over time. In order to get the most advantage for estimation out of the little amount of data available over time, it is very important to make the most efficient use of the data across economies to estimate the overall variance of the behavioural relationship containing variables that differ substantially from one economy to another "between variance", and variability within KBE units over time "within variability". Hence, it is suggested to run the "Hausman" specification test, which is usually used to examine the significance of the difference between using the fixed effects and the random effects estimates (Hsiao, 2003). This can be hypothesised as follows:

- ***H0*** : No correlation between the random effects and the regressors. Accepting this allows us to use the random effects model.
- ***H1***: There is a correlation between the random effects and the regressors.

The Hausman test results for IMD-WCY index as a dependant variable are presented in Table 7.1. The differences between the two sets of estimates are tested as a block using χ^2 or Chi-square. The χ^2 with 4 degree of freedom between all independent coefficients has produced $\chi^2 = 105.86$ with probability value of 0.00. These results suggest the null hypothesis of no correlation between the independent variables and the random effects model should be rejected. This process is repeated to test all TSCS models, and similar decisions are concluded. Therefore, all TSCS regression are carried out using the fixed effect model only.

Panel data analysis suffer from some draw backs such as heterogeneity and selectivity biases. Panel data analysis produce one average estimate for all entities over time regardless of the differences between different nations, and it does not randomly select the samples from the full data set presented (Hsiao, 2003).

7.2.1 Artificial Neural Network Techniques

In many practical situations, forecasting with short time periods or missing values is highly desirable. Some contributions are introduced to deal with such cases, including the use of SOM, higher frequency data, the Bayesian approach, and the meta-analysis to forecast with little data (Meade and Islam, 2006). The collected data can be dealt with as a time-series alone, but it does not have enough time periods. Only three to five years period data across the selected indicators are available, which is very little to give a meaningful forecast to the change over time and it is short of the full economic prediction cycle threshold which usually requires data for more than seven years (Korotayev and Tsirel, 2010).

It is established that ANN can be used to forecast short or limited data set, and it can predict linear and non-linear relationship (Adeodato et al., 2009; Ilonen et al., 2006; Crone et al., 2011). ANN models outperformed several traditional statistical techniques, since these techniques are essentially linear techniques and they require long time-series to be able to predict successfully. However, building an ANN for forecasting with short time periods is not a straight forward task, because many things must be taken into consideration to get the desired accuracy, such as the ANN type, the layers counts, the counts of hidden neurons in

Table 7.1: Hausman test for IMD-WCY index: fixed vs. random effects.

Test Summary		Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Cross-section random		105.8656	4	0.0000
Cross-section random effects test comparisons:				
Variable	Fixed	Random	Var(Diff.)	Prob.
ITU	1.47068	0.36986	0.00019	0.0000
WB	-10.29712	-1.37133	0.01278	0.0000
WEF	1.03969	0.68341	0.00016	0.0000
INS	-0.03584	-0.00603	0.000001	0.0000

each layer, the training method, the activation function, data preparations and divisions etc. (Wilson et al., 2002).

Artificial Neural Networks are becoming the trend for their precision in predictions, clustering, modelling and trend analysis. Some techniques are more popular than others, and the ANN are now considered to be the most popular fitting tool with high predictive accuracy compared with other computational intelligence methods and of course compared to the traditional linear statistical methods. ANNs are an extension to the least square method, but it is considered as an adaptable non-linear regression (Baker, 1999).

In many studies, it has been shown that ANN can model any functional linear and non-linear relationship, and that such models are better than regression since regression is essentially a linear technique used to solve non-linear problems (Wilson et al., 2002). ANNs are commonly categorized in terms of their corresponding training algorithms; mainly supervised and unsupervised training. Back propagation is a common method of training feed-forward ANNs. Feed-forward neural network is usually used for applications that require fitting a set of inputs to a particular targeted outputs. In contrast, a self-organizing map (SOM) or self-organizing feature map (SOFM) is a type of ANN that is trained using unsupervised learning to produce a low-dimensional, discretized representation of the input space of the training samples, called a map.

7.2.1.1 ANN Training Techniques

Training feed-forward neural network usually happens in three steps: each input $X_1, ..., X_n$ will be feed-forwarded to train the network to capture the data pattern then it sends the signal of this pattern to the hidden neurons. The hidden neurons $1, ..., h$ computes the activation function using either the binary sigmoid function or the bipolar sigmoid function, and sends the results to the output Y_h. Using the gradient descent method, the error generated will be back-propagated after minimising the sum squared error of the outputs against the specified targets. The network keeps cycling through the entire set of training vectors (each complete cycle is called an epoch) and keeps adjusting and updating the connection weights W_{hn} accordingly, till it reaches the least possible error or what is known as the global minima (Fausett, 1994).

SOM on the other hand is a suitable technique for multidimensional data clustering (Ilonen et al., 2006) and it helps to avoid the curse of dimensionality and the heterogeneity bias which arise among cross-sectional units (Hsiao, 2003). Thus, the aim is to find similar countries (not necessarily territorial neighbours) by using comparative measures and sensitivity analysis. This can be achieved arithmetically by computing the first and the second derivative for each economy along the time series (Zimmermann, 2011). Alternatively this can be achieved manually by grouping similar countries based on their similar scores and observing the moving ones. The task becomes daunting when presented with many countries, many reported scores with high variations between their reported scores. If they are similar and should be classified as neighbours, so an automated method to find similar competitive KBE within multi-dimensional data set would be highly desirable.

Although SOM algorithm is a well established non-linear mapping tool, and it has many beneficial properties, such as tolerance for incomplete and small data set (Ghaseminezhad and Karami, 2011), however, few issues need to be tackled to get efficient results. It is suggested by Thang et al. (2003) to train the SOM in two phases: ordering phase and then tuning phase. The ordering phase helps the network to quickly scan large area in search of related neurons, and not getting stuck in a local minima. The ordering phase usually requires setting high learning rate, large distance and small number of epochs. The tuning phase requires higher number of epochs, small distance and low learning rate to tune-up the rough structure of the earlier phase to produce a well organised and tightly coupled map. It is also suggested to use some heuristics measures to evaluate the efficiency of the trained SOM by measuring and comparing the quantization error (QE) and the topographic error (TE), eventually aiming for a low TE and QE.

7.2.1.2 Data Division for Training and Testing

Collected data from complete 51 economies was used as inputs to the feed-forward neural network to predict the future values for five indices. This simple mode of

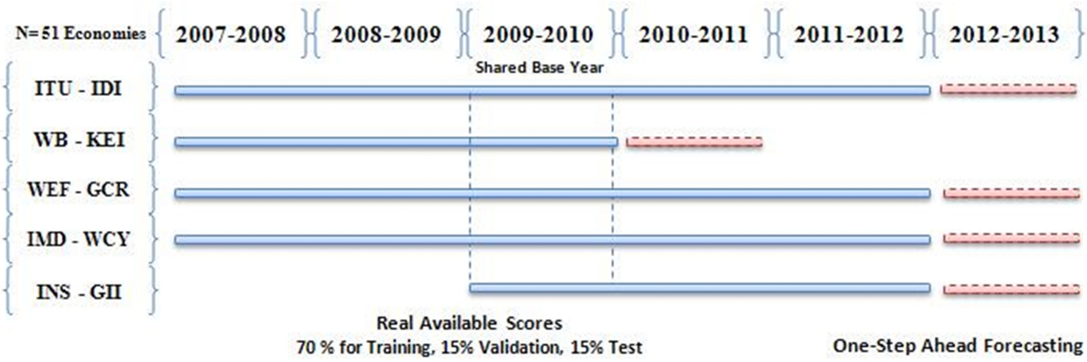


Figure 7.2: Diagram of data availability, inputs and division for ANN one-step ahead forecasting.

prediction was later modified by inputting the ANN with balanced panel data structure. The simple ANN mode was referred to as "ANN I" and the modified version as "ANN II", as depicted earlier in Figure 7.1. For both modes, input data were divided randomly into a training, validation and test subsets in the following fashion 70%, 15% and 15% respectively, for weight learning, over-fitting prevention and performance validation (Adeodato et al., 2009). The network was trained using the Levenberg Marquardt backpropagation algorithm, which is known as one of the most effecient ANN training algorithem (Kershaw and Rossini, 1999) (Atiya et al., 1999). Figure 7.2 shows used data availability and division for ANN forecasting.

7.2.2 KBE Forecasting Using ANN with Panel Data Structure

The modified ANN forecasting model, ANN II, integrate supervised with unsupervised learning using feed-forward neural network and SOM. A schematic diagram of the proposed model is presented in Figure 7.3, which consists of two main units; a three-layers feed-forward neural network and p processing neurons SOM. The purpose of the first unit is to learn and predict the scores of nations using past observed data. SOM is made of input nodes and a competitive layer or output layer of neurons where the groupings are taking place.

The feed-forward neural network with only five to eight hidden neurons using

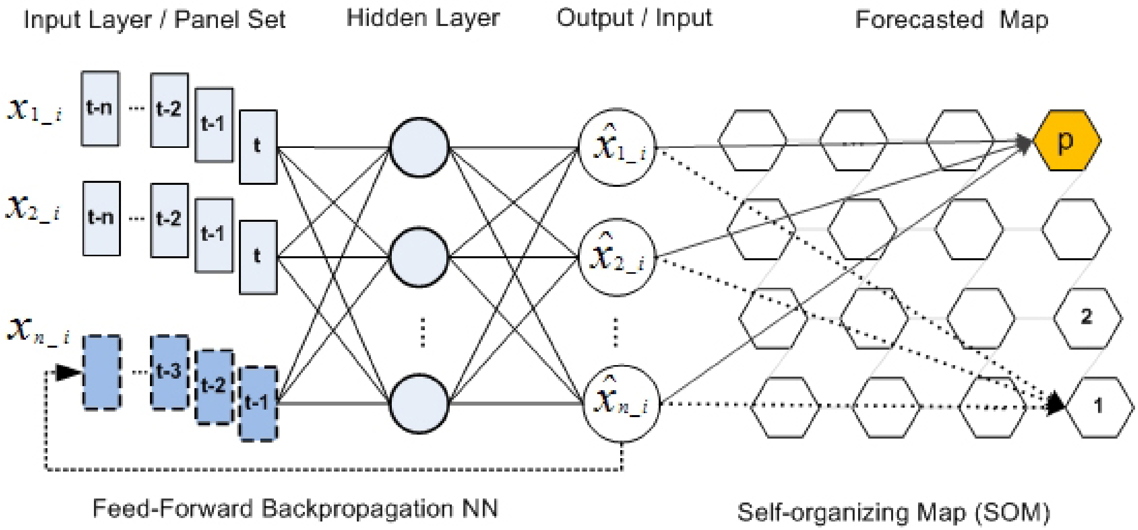


Figure 7.3: Proposed KBE forecasting model.

tansig transfer function and linear output neurons is used. The results obtained were then compared with two types of regressions: Panel data, time-series cross sectional and linear multiple regression. The aim is to allow the network to map between the cross-sectional inputs (independent variables) and the cross-sectional targets (dependent variables), to produce an accurate prediction of output measures. With the time-series part of the panel data the mission is to allow the feed-forward neural network to learn the pattern from the historical time-series observations on each of the cross-sectional selected economy to forecast the future KBE score for any nation.

For SOM network, the input consist of six indicators for the forecasted scores for the 51 selected economies, where they are connected to each neuron in the array. Since knowledge progress and competitiveness scores are available for some countries and reported by more than one index, SOM is employed to combine and cluster using all indices at once as a unified KBE index to capture the homogeneous, the progressing or accelerating KBEs.

Usually, SOM learns to classify input nodes according to how they are grouped in the input space, therefore, it could recognise countries according to their forecasted score and it would organise those with similar scores and show them as neighbours even if they are geographically not. Thus, SOMs learn both the dis-

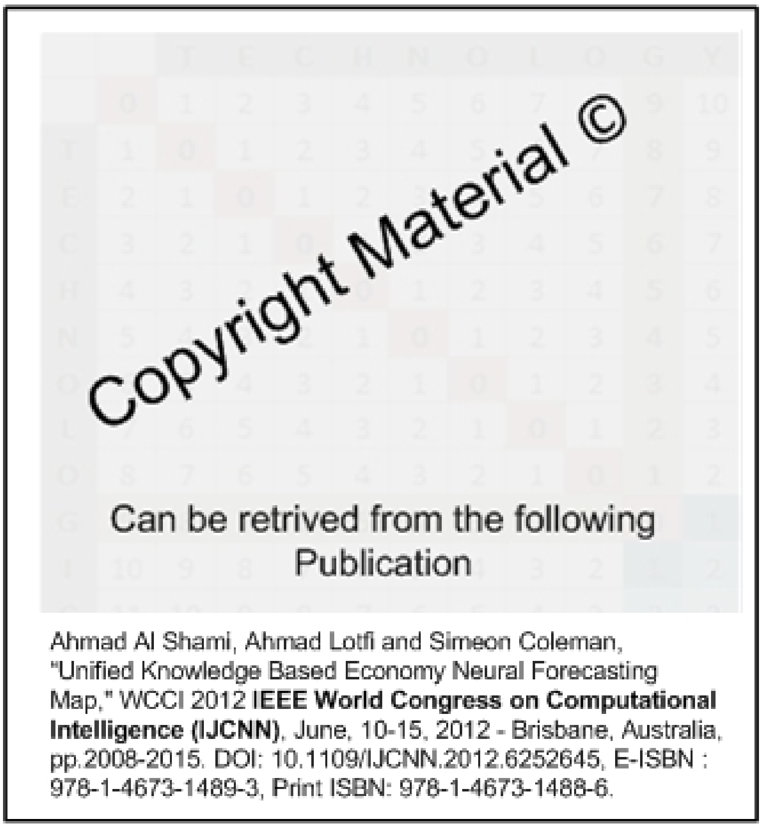


Figure 7.4: SOM map before and after using PCA.

tribution and topology of the input nodes they are trained on. Since there is no strict scientific method to determine the right size for the SOM map, different values were experimented with, until satisfied with the results using a 5 by 3 layer of neurons to produce $p = 15$ different meaningful classification for KBE scores from all indicators. Also, this study is proposing the use of PCA before using SOM to cluster the forecasted scores to produce the suggested UKFM.

To illustrate the benefits of applying PCA filtration before training the SOM map, In Figure 7.4 plotted the forecasted scores before and after training the SOM neurons, to show how SOM dealt with the raw data (in the upper or “before” graph) vs. the (lower or “after”) applying PCA, Variance Accounted For (VAF) technique (SAS.com, 2011), which helped retaining only the components that accounted for more than 5% of the total variance in the dataset.

The “before” map graph shows how SOM has missed some data points (pointed to by the arrows) and in one case it placed the cluster in odd location - far off from the rest of its neighbours - just to account for such outlier point. On the other hand the lower or "after" applying the Principle Component VAF technique, the SOM started to offer a better self-organize, well-rounded coverage and consistent clustered groups, so that each neuron organises a different groups of the input space, and related neurons got better connected to the nearby neighbouring group.

7.3 Performance, Analysis and Results

In this section the results for indicators prediction and mapping economies are presented. Accuracy and performance of the results are measured using measured defined below.

7.3.1 Accuracy and Performance Measures

The accuracy and performance for the methods utilised in this study were tested using both the Mean Absolute Percent Error (MAPE) and the Mean Square Error (MSE); MAPE can be expressed as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{x_t - \hat{x}_t}{x_t}\right| \tag{7.2}$$

while

$$MSE = \frac{\sum_{t=1}^{n}(x_t - \hat{x}_t)^2}{n} \tag{7.3}$$

where the smaller the MAPE or MSE value the better the model.

7.3.2 The Forecasting Performance Results

After running all experiments using all methods mentioned earlier across and between different indicators, many combination results are generated. Table 7.2 shows the overall MAPE and MSE performances results for all used techniques.

The detailed MSE performance results for both ANN networks with cross validation step, the mean and standard deviation after 10 runs are also listed.

The prediction results suggests that all models performed well, except the linear regression. It is noticed that the performance of the models depends on the indicator that tries to predict. For example it is noticed that WB-KEI can perfectly fit for ITU-IDI but not the other way around. This is due to the fact that ITU-IDI constituent elements are used by the WB-KEI under its Information, Communication and Technology (ICT) pillar. In conclusion by structuring and feeding balanced, normalised panel data to the ANN model (ANN II), a slightly better predicted results are obtained compared with the ANN without panel data structure (ANN I).

Table 7.2: Forecast overall performance results for the selected KBE indicators.

MAPE						
Method		WEF-GCI	WB-KEI	ITU-IDI	IMD-WCY	INS-GII
LMREG		8.603	21.930	13.532	24.741	33.263
TSCS		3.505	3.542	4.441	8.760	20.521
ANN I		1.86	1.94	2.10	3.31	5.32
ANN II		**1.12**	**1.12**	**1.20**	**1.31**	**3.02**
MSE						
		WEF-GCI	WB-KEI	ITU-IDI	IMD-WCY	INS-GII
LMREG		0.241	0.643	0.582	0.933	0.651
TSCS		0.104	0.104	0.240	0.150	0.305
ANN I	**(10 runs with cross validation)**					
Test-MSE	Minimum	**0.030**	**0.035**	**0.059**	**0.164**	**0.279**
	Mean	0.091	0.245	0.279	0.210	0.782
	Std.Dev.	0.074	0.064	0.059	0.037	0.081
Train-MSE	Minimum	2.714e-04	2.98e-03	7.17e-3	1.358e-03	0.189
	Mean	1.114e-02	4.58e-03	7.17e-2	1.413e-02	0.166
	Std.Dev.	0.041	0.051	0.029	0.071	0.090
ANN II	**(10 runs with cross validation)**					
Test-MSE	Minimum	**0.021**	**0.044**	**0.050**	**0.030**	**0.242**
	Mean	0.0851	0.128	0.137	0.160	0.662
	Std.Dev.	0.012	0.016	0.018	0.034	0.094
Train-MSE	Minimum	1.131e-07	2.58e-04	9.17e-26	2.413e-04	0.110
	Mean	2.731e-04	4.81e-03	6.67e-9	7.39e-03	0.180
	Std.Dev.	0.013	0.016	0.020	0.037	0.042

Bold values signify higher and more accurate performance.

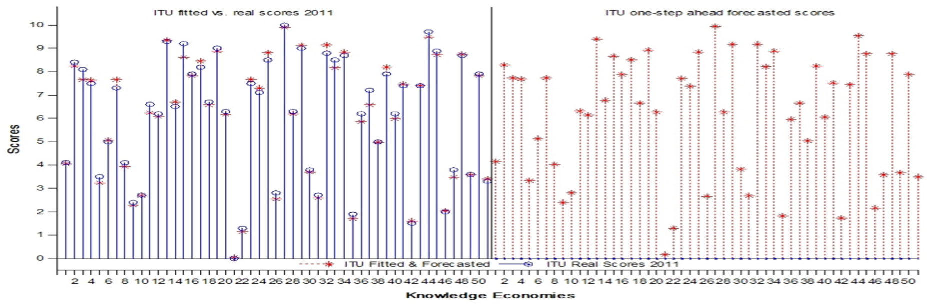
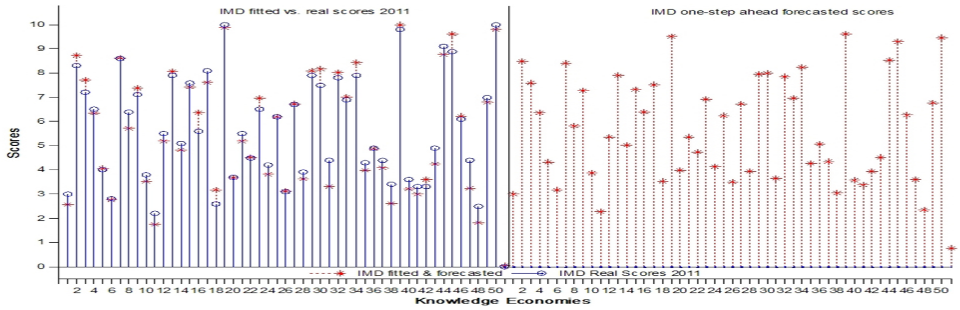
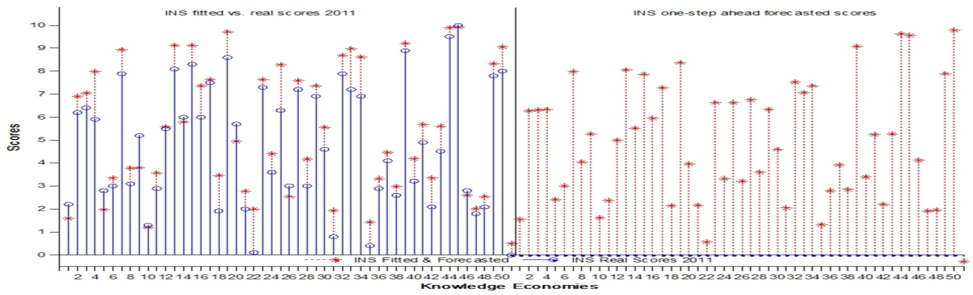


Figure 7.5: Predicted vs. real scores and one-step ahead forecasting based on the proposed ANN II for ITU index.

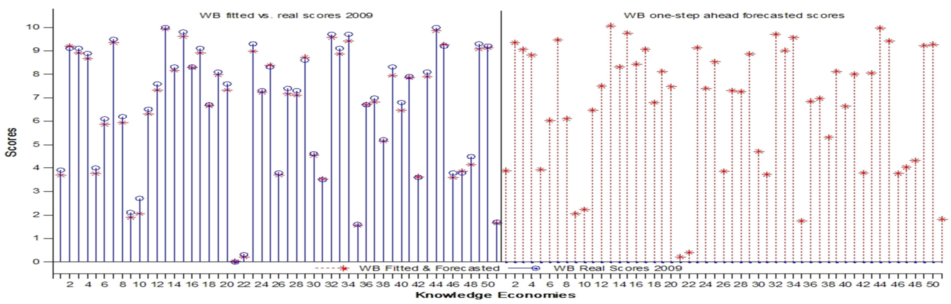
Figure 7.5 shows results obtained by ANN II for fitted and forecasted scores results for ITU index for 51 economies. The fitted results for the last reported year appears in the left hand side, and the one-step ahead forecasted scores are located to the right. One can obviously notice the near “exact fit” between the ANN II predicted results vs. real scores, which gives a large confidence in the obtained forecasted results. List of countries with their numerical labels and the abbreviations are provided in Table 7.3. Figure 7.6 shows the predicted and forecasted results obtained by ANN II for IMD, INS, WB and WEF indices.



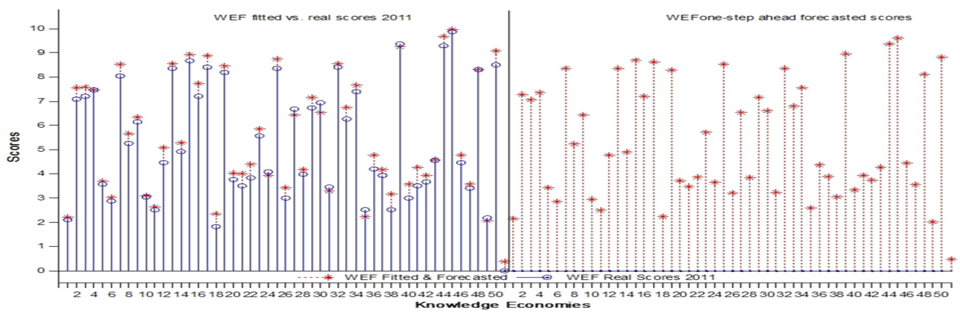
(a) IMD-WCY



(b) INS-GII



(c) WB-KEI



(d) WEF-GCI

Figure 7.6: Predicted vs. real scores and one-step ahead forecasting based on the proposed ANN II for IMD, INS, WB and WEF indices.

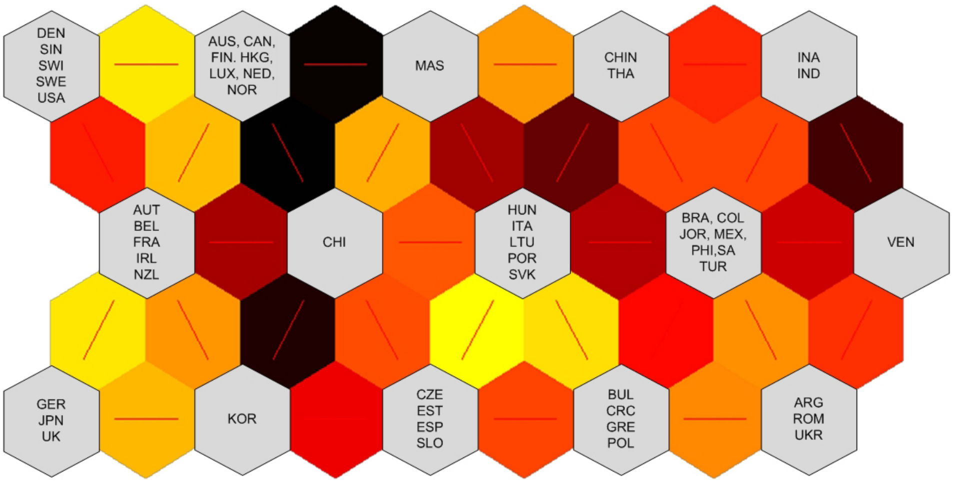
Table 7.3: Labels, abbreviations and name of economies used in this study.

NO.	Eco.	Economy	NO.	Eco.	Economy	NO.	Eco.	Economy
1	ARG	Argentina	18	GRE	Greece	35	PHI	Philippines
2	AUS	Australia	19	HKG	Hong Kong	36	POL	Poland
3	AUT	Austria	20	HUN	Hungary	37	POR	Portugal
4	BEL	Belgium	21	IND	India	38	ROM	Romania
5	BRA	Brazil	22	INA	Indonesia	39	SIN	Singapore
6	BUL	Bulgaria	23	IRL	Ireland	40	SVK	Slovak Republic
7	CAN	Canada	24	ITA	Italy	41	SLO	Slovenia
8	CHI	Chile	25	JPN	Japan	42	SA	South Africa
9	CHN	China	26	JOR	Jordan	43	ESP	Spain
10	COL	Colombia	27	KOR	Korea (Rep.)	44	SUI	Sweden
11	CRC	Croatia	28	LTU	Lithuania	45	SWE	Switzerland
12	CZE	Czech Republic	29	LUX	Luxembourg	46	THA	Thailand
13	DEN	Denmark	30	MAS	Malaysia	47	TUR	Turkey
14	EST	Estonia	31	MEX	Mexico	48	UKR	Ukraine
15	FIN	Finland	32	NED	Netherlands	49	UK	United Kingdom
16	FRA	France	33	NZL	New Zealand	50	USA	USA
17	GER	Germany	34	NOR	Norway	51	VEN	Venezuela

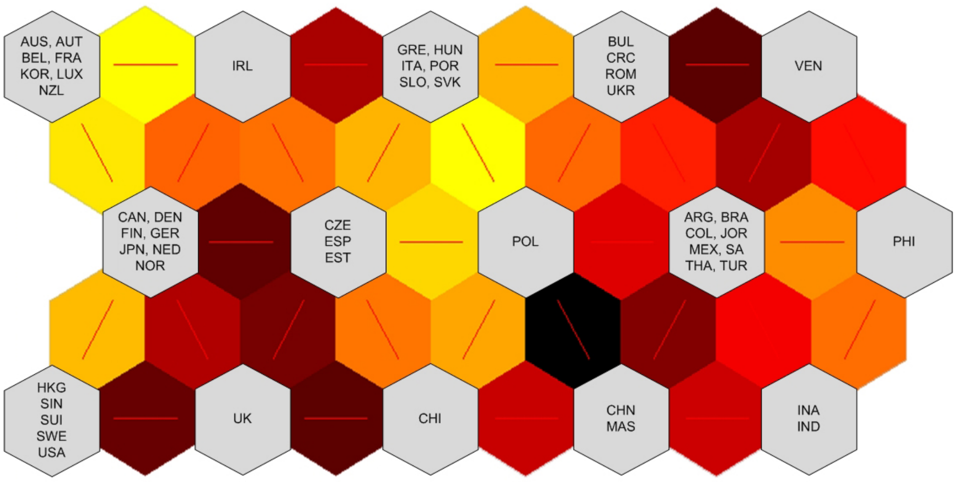
7.3.3 The SOM Mapping Results

The final step is to visualise and evaluate the forecasted results obtained by ANN II in a user friendly fashion. This will help decision makers to focus on the results rather than the process. A series of the forecasted results can be produced as a result of the SOM clustering model depicted in Figure 7.3. The coloured SOM maps presented in Figure 7.7(a) and Figure 7.7(b) respectively to show the change in the state of knowledge based economies from 2009 to the one-step ahead forecasting results for the selected 51 economies, which are being clustered into 15 groups, where each neuron or group contain the clustered homogeneous KBE members.

The SOM map has the following shapes and colour codes: The grey labelled hexagons represent the neurons or the 15 different clustered groups, where each contains the "neighbours" KBEs. The hexagons connected by lines, and the colours in the regions containing the lines indicate the distances between neurons. The darker colours represent larger distances or less related KBEs. The lighter colours represent smaller distances or closely related groups of KBEs. The order of the clustered groups goes as follows: Highly competitive KBEs are placed on the left hand side, less competitive groups are gradually placed to the right.



(a) SOM for 2009 real scores.

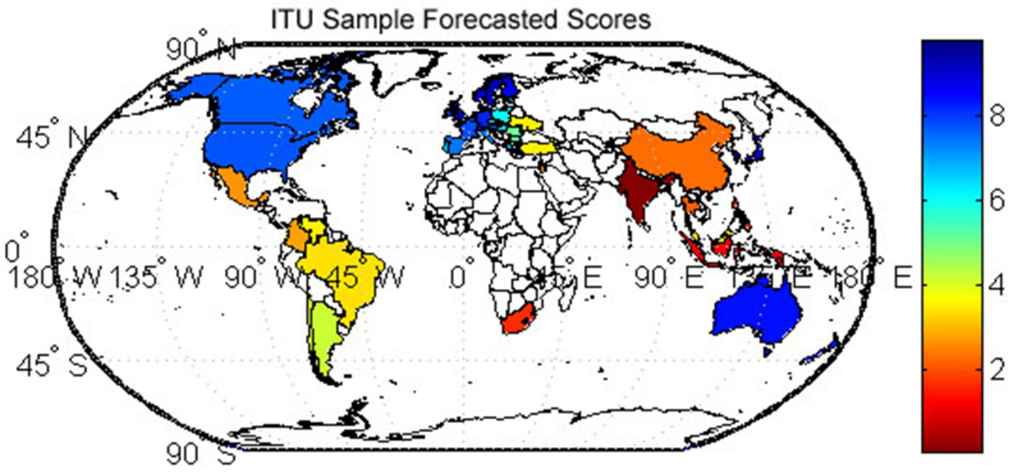


(b) SOM for 2012-2013 forecasted scores.

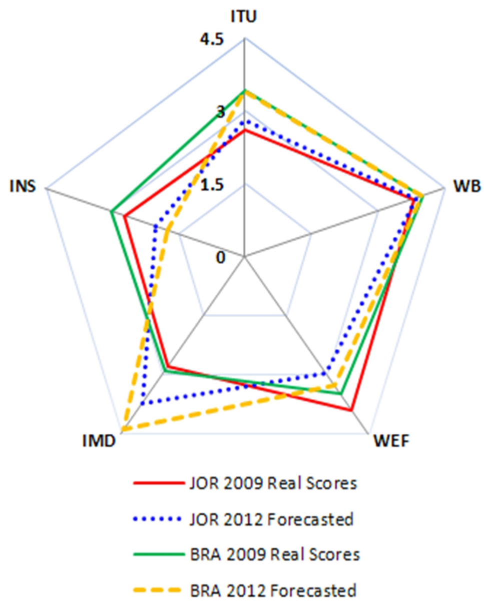
Figure 7.7: SOM clustering map results form selected KBE indicators.

Even with a casual look over the produced SOM maps, one can clearly notice some obvious results on the left hand side of the graph , such as the USA, Denmark, Hong Kong, Sweden, Singapore and Switzerland, being shown in one cluster as very close neighbours. Canada, Austria, Finland, Luxembourg, Netherlands and Norway are highly related KBEs and located near each other. These results are anticipated and resembles the famous G7 grouping, except for Italy which the SOM placed in a separate cell and far from the advanced nations as it is facing many challenges nowadays. What is new is to see less directly understandable similarities for other countries, such as Brazil, Turkey, Mexico, South Africa, and Jordan shown as neighbours, even though these KBEs are geographically distant apart.This result resembles the famous BRICS and MIKT groups. SOM clustered Brazil and South Africa in one cell which is tightly coupled to India in a neighbouring cell and linked to China in another cell but within close proximity. China and India are in separate cells but with close and strong distant link. Mexico and Turkey are also in one cell group with Brazil and South Africa. However, Korea is extremely far from all as being ranked between the top ten KBEs in the world. One can infer many similar results by comparing SOM grouping to other famous grouped economy such as PIGS (Portugal, Italy, Greece and Spain) or PIIGS(Portugal, Italy, Ireland, Greece and Spain)...etc.

By comparing the 2009 real scores SOM map against the forecasted SOM some remarkable facts were revealed about the positions of the countries: the positions of countries are stable with few exceptions lead by the most active KBEs such as Korea, Malaysia, China, Argentina, and India. As one would expect these economies to be in close distance cluster, but this difference in grouping between these countries proves the volatility and the nature of these fast growing knowledge economies. This result draws the facts which is due to the low scores obtained by India and China for ICT use, skills and access, and the opposite was true for Korea, Malaysia and Argentina. Another interesting fact reviled was that “size does not matter” when it comes to the most competitive knowledge economies in the world, which as the SOM map shows is led by small economies like: Singapore, Switzerland, Sweden and followed by another small neighbours economies such as Denmark, Finland, Netherlands and Norway. This could form a new group to be called “DFNNSSS” to abbreviate the Denmark, Finland, Netherlands, Norway, Switzerland, Sweden and Singapore. This fact result should give real hope to small developing economies.



(a) ITU 2012-2013 forecasted scores using Global Choropleth Map



(b) Jordan vs. Brazil 2009 real scores and 2012-2013 forecasted scores

Figure 7.8: Visualisation techniques for one-step ahead forecasted and clustered results.

On the other hand if evaluating the results of a single indicator or a specific or few KBEs is the main concern, additional visualization tools, such as the

Choropleth map and/or the Radar chart could be useful. Figure 7.8(a) show a scaled thematic world map where the ITU(IDI) for one-step ahead forecasted scores for the sampled 51 economies are presented. The Choropleth map usually provides a quick glimpse of how the forecasted scores are distributed across the globe. In Figure 7.8(b) where data for two economies from the same cluster were plotted and visualised for Jordan versus Brazil using ITU, WB, WEF, IMD and INS scores for the one-step ahead forecast and the 2009 real score, as an example for further and more detailed possible visualisation. The radar chart could also be useful for validation of the accuracy of the SOM clustered results, thus evaluation and/or prediction of group specific competitiveness of countries in time (dynamic process). Therefore, this radar chart infers a successful, valid and accurate SOM classification of the homogeneous yet of different size and geographically distant KBEs.

7.4 Summary

Studies comparing panel data and computational intelligence techniques are sparse. This chapter makes four important contributions. First, it compares three different predictions methods to reveal the performance superiority between them; the ANN was a clear winner, as it far exceeded the linear multiple regression and the famous TSCS regression. Second, it proposed a new approach on feeding the ANN with balanced, normalised panel data, which produced a slightly better performance than feeding the network with just normalised scores. Third, the method overcome the limitations of short time periods and it managed to forecast using the ANN to produce one-step ahead future scores for six major knowledge competitiveness indicators, including not reported yet indicator since 2009 like the KEI from the World Bank. Finally, it employs SOM to unify and cluster the forecasted results to automatically visualise and evaluate the results in a user friendly fashion, to capture homogeneous, active or accelerating knowledge economies.

Chapter 8

Conclusions and Future Works

8.1 Thesis Summary

The work presented in this thesis is an attempt to answer the research question both from the practical and theoretical point of view. Based on the results obtained from this research, it can be concluded that computational intelligence techniques alone or combined with statistical means, can provide novel way to build future composite indicators. The proposed approach works better for measuring non-linear and unpredictable phenomena such as macro-knowledge, competitiveness, or sustainability issues. Specifically, we focused our research for developing a unified macro-knowledge measure suitable for any nation, but will mostly benefit developing or underdeveloped countries, where the data is usually missing, and the near future behaviour of the economy is unpredictable.

The aim of the research was to investigate the use of computational intelligence for efficient knowledge mining, missing data imputation, variables weighting, aggregation and forecasting, for the purpose of producing a unified macro-knowledge competitiveness indicator. We have investigated a wide array of methods for qualitative or text based variables matching; quantitative variables weighting, aggregating, forecasting and clustering. The research was conducted to overcome the pitfalls of already existed methods for constructing synthetic composite indicators, and to show that CI can be used to construct a data driven, systematic, non-bias, simple yet intelligent SCI. This new breed is used in this research to measure, visualise, identify or evaluate stable, progressing or accelerating KBEs.

Data provided for the investigation were mined and collected from various reputable international organisation, as well as from prestigious academic institutions. For analysis, tests and comparison purposes, data were organised and reported as three different empirical case studies. First, the final or overall scores from few indicators, were filtered of anomalies such as missing data and outliers, then treated using multivariate analysis and principle component analysis. This was done to initially foresee the possibilities of conducting the initial aim and objectives. Second, collect to include/exclude qualitative data using fuzzy knowledge mining techniques, to build a data driven, intelligent "learning" qualitative taxonomy, impute missing data, weight and aggregate using purely computational intelligence techniques, which outperformed statistical methods. Third, final existing scores with temporal data were collected for the purpose of devising a suitable KBE prediction and forecasting techniques. Different approaches for data visualisation and compression were investigated. Different robustness techniques were also employed to test the validity and strengths of the proposed and used methods. ANN and specifically SOM were used to predict, forecast and cluster the behavioural patterns of KBEs progress.

In summary, throughout this research, original knowledge on composite indicators, knowledge mining , data analysis, missing data, weighting, aggregation and forecasting a vast amounts of non-linear data via Computational Intelligence techniques has been obtained. In the remaining part of this chapter, the research conclusions with critical assessment, reflections and suggestions for future research are presented.

8.2 Concluding Remarks

This thesis attempts to provide a knowledge management system based on pure CI methods. The developed system is capable of evaluating, measuring, describing, forecasting and analysing the main issues that affect knowledge economies on a macro and micro levels. Conclusions for the main features of the project are summed.

8.2.1 Intelligent Qualitative Taxonomy

This thesis highlights the need for a flexible and intelligent qualitative data collection and representation techniques. Fuzzy Proximity Knowledge Mining (FPKM) techniques are used and tested using real qualitative variables to establish a robust and intelligent way to build future qualitative taxonomies. The suggested FPKM consists of two major steps: focused web mining and fuzzy text matching such as the the Wagner-Fischer dynamic programming algorithm for computing the Levenshtein distance. The results of these concepts of mining useful and pre-existing knowledge to make "meaning-driven" qualitative taxonomy for a new indicator proofed useful, without the need for any "experts" interference or biases. Moreover, various web mining, exact and fuzzy text matching techniques are used to detected, include/exclude and group relevant variables for the purpose of developing the theoretical framework for any brand new SCI.

8.2.2 Intelligent Synthetic Composite Indicators (*iSCI*)

In this study the main aim was to introduce a new way of developing a new type of composite indicators that would overcome the subjectivity of "experts" and the pitfalls of the statistically based SCI by using computational intelligence means. As a direct result of this research, a good understanding of the different statistical and CI techniques to build SCI are identified. This study has specifically investigated the use of CI techniques such as Fuzzy c-Means (FCM) and Vector Quantisation clustering to identify natural aggregation in quantitative datasets to allow for concise representation of the relationships embedded within the variables and to generate the final *UKCI* composite indicator. Different methods were investigated for the purpose of missing data, weighting, aggregations of variables into a smaller subset while avoiding any organisation's or expert's subjectivities or opinion biases. FCM and its derived strategies, specifically, the Optimal Completion Strategy (OCS) and the Nearest Prototype Strategy (NPS) are also investigated for missing variables scores.

The results of this investigation show that FCM is better than VQ and other statistical based distance measures for SCI developments. FCM represented by its iterative process to learn the entropy of the inputs variables and present the needed clusters. However, FCM is usually affected by the presence of outliers and the low dimensionality of the data which made it rather difficult to identify the anomalies and outliers within the input data. To tackle the high dimensionality of the data, to identify the outliers and anomalies of KBEs progress patterns and to avoid any positive bias couple of sound experiments were conducted (Pearson Correlation Coefficient , PCA analysis, outliers detection and extensive knowledge on the nature of the inputs).

The proposed FCM based framework is successfully able to identify a cluster centre, which is weighted by the mean of all points based on their degree of belonging to the cluster to represent the natural aggregation score for any input data sets, hence the desired SCI. This novel and concise representation of the relationships embedded within the variables lunched the 3rd generation of composite indicators, branded as The Intelligent Synthetic Composite Indicators (*iSCI*). The proposed intelligent indicators carry special character as they are data-driven and therefore such indicators have the capabilities to accurately rank nations based on non-bias “learning”.

The validity and robustness of all techniques are evaluated using Monte Carlo simulation. The results obtained suggest an alternative novel, intelligent and non-biased method of building future composite indicators using purely CI techniques. Additionally, the results and knowledge gained from of this research could be extended for other disciplines such as decision science, financial analysis and portfolio managements etc.

8.2.3 Application of *iSCI* for Measuring Macro-Knowledge Competitiveness

The FCM concept was first used to present a unified ICT index as an empirical case study. The proposed construction methods proofed valid and robust to launch the 3rd generation of SCI, referred to as the intelligent SCI or *iSCI* for short and throughout this thesis. The *iSCI* is then applied to build the suggested *UKCI* baskets, themes and the final scores for Fifty Seven knowledge based economies initially and extended to include the MENA region KBEs. In total Seventy Three KBEs are included in the *UKCI* analysis.

The *UKCI* is the answer to the growing need for an 'all inclusive' unified indicator that encapsulates the major macro-knowledge indices. This is in line with one of the main objectives of this thesis for "Index of Indexes" on KBE competitiveness. The *UKCI* encapsulates the strengths of seven major KBE and competitiveness indicators with complex and multi-dimensional nature, into a single meaningful value. It also avoids the weaknesses of these indicators in such that it consists of sub-units to allow for top-down, precise diagnostics, interpretability, diffusion and monitoring of macro-knowledge progress.

8.2.4 Macro-Knowledge Competitiveness Prediction and Forecasting

This thesis also shows that the KBE competitiveness and progress can be efficiently predicted and forecasted using ANNs and advance econometrics techniques such as Panel Data Analysis.

Three different forecasting methods including Panel Data: time-series cross sectional (TSCS), Linear Multiple Regression (LMREG), and Artificial Neural Network (ANN) were investigated. The ANN forecasting model outperformed the TSCS and LMREG. The proposed KBE forecasting model utilizes 2-stages ANN models which is fed with panel data set structure. The first stage of the model consists of a feed-forward neural network that feeds to a Kohonen's Self-Organizing Map (SOM) in the second stage of the model. A feed-forward neural network is used to learn and predict the scores of nations using past observed data. SOM performance was also improved through using PCA to act as a filter before feeding SOM the forecasted scores. This novel process was able to produce a Unified Knowledge Economy Forecast Map (UKFM) which reflects the overall position of homogeneous KBEs, and it can be used to visualise, identify or evaluate stable, progressing or accelerating KBEs.

The proposed models and the forecasted results were verified tested and compared to show the model ability and accuracy in one step prediction. The output of the UKFM could be used to forecast and visually combine scores for any given economy especially for developing economies where the scores usually missing or not reported by one or more of the used indicators.

In general, the major findings of this work in terms of macro-knowledge competitiveness prediction and forecasting with limited time series dataset, collected from internationally recognised organisations including the World Bank and the International Telecommunication Union among others, are listed below:

- Studies comparing panel data and computational intelligence techniques are sparse. Specifically this was true for analysis to overcome the limitations of short time periods, representing data collected from internationally recognised organisations such as the World Bank, The World Economic Forum and few others to produce one-step ahead future scores.
- The produced results indicate that when feeding panel data to a neural network algorithms it outperformed traditional ANN and statistical time series prediction methods. Additionally, the results presented in this research show that combining TSCS and ANN are very promising approach for datasets collected for many entities over short or limited time periods.
- SOM could be used to unify and cluster the forecasted results for six complex, multi-dimensional macro-knowledge composite indicators. The results could be presented in multi grid maps to automatically visualise and evaluate the results in a user friendly fashion in order to capture homogeneous, active or accelerating knowledge economies.
- Size and locations does not matter when it comes to competitive KBEs.

Overall, this research could play a vital role in the field of Decision Science and neuro-fuzzy economics. The results obtained suggest a novel, intelligent and non-biassed way compared to traditional or statistical methods when building, not only the presented *UKCI*, but for developing any future composite indicator for any other fields.

8.3 Future Works

Further investigation, in which future works could proceed, are listed below:

- A further direction for investigation is to keep track of the used composite indicators over a longer period of time. A five to seven years data set would be necessary for a more reliable KBE competitiveness and progress forecasted results. Some baskets could be very hard to predict even with the presence of long time periods. For example predicting the economic progress of politically unstable country. This task is complex, as not only must monitor the country in a continuous basis, but it must also predict whether such country has a potential to erupt in a coup, riots or a sudden revolution after decades of fixed status.

- This study will serve as a basis for future studies in which different CI techniques could be tested. For example, the current study has only examined Type 1 Fuzzy Logic. A future study investigating the impact of using Type 2 Fuzzy Logic (Karnik et al., 1999; Mendel and John, 2002; Mendel et al., 2006; Coupland and John, 2008) to increase the accuracy of the presented models. For example the conventional FCM algorithm used in this study to develop the *iSCI* could be replaced by type-2 FCM algorithm (Rhee and Hwang, 2001; Hwang and Rhee, 2007), which may converge to a more precise centroid location when using cluster centres.

- It would be appealing to build a user friendly on-line dashboard system to allow for interactive and immediate analysis results of the presented *UKCI* or similar application.

- It will be appealing to extend the work to develop the semantic structure of the behaviour of a KBE where the predicted values are communicated with the decision makers in linguistic terms. This will be achieved by taking the results of this thesis and feeding them to an ontology editor such as Protégé, Knoodl or other similar software.

8.4 Problems Confronted

Few problems were met while developing the *UKCI* indicator. The major problem was data availability, particularly in the innovation and public administrative as well as in the intellectual capital areas, where these concepts are fairly new and organisations started the collection of data since 2008. Because the data acquired was for a short time period and with many missing data points, the present research presented some ideal solutions to deal with such shortcomings, in hope that more data would be available as the time goes by.

Another problem which is inherited in an all-inclusive approaches to measuring KBE, and consequently the *UKCI* indicator, rest in the fact that knowledge economy indicators tap slightly different variations in the underlying constructs, resulting in some of the variables may possibly measuring the same thing. To combat this in a non-biassed way, a novel qualitative taxonomy based on fuzzy proximity knowledge mining was devised to identify and naturally group the relevant variables.

Appendix A

The following is the pseudo code of LBG VQ algorithm (Linde et al., 1980):-

1. Given T Fixed $\epsilon > 0$ to be a "small" number.
2. Let $N = 1$ and

$$C_1^* = \frac{1}{M} \sum_{m=1}^{M} X_m. \tag{1}$$

   Calculate

$$D_{avg}^* = \frac{1}{Mk} \sum_{m=1}^{M} \|X_m - C_1^*\|^2 . \tag{2}$$

3. Splitting: For $i = 1, 2, ..., N$,

   set

$$C_i^{(0)} = (1 + \epsilon) C_i^*, \tag{3}$$

$$C_{N+i}^{(0)} = (1 + \epsilon) C_i^*, \tag{4}$$

   Set $N = 2N$.

4. Iteration: Let $D_{avg}^{(0)} = D_{avg}^*$. Set the iteration index $i = 0$.

i- For $m = 1, 2, , M$ find the minimum value of

$$\left\| X_M - C_n^{(i)} \right\| overalln = 1, 2, , N. \tag{5}$$

Let n^* be the index which achieves the minimum.

Set

$$Q\left(X_m\right) = C_{n^*}^{i}. \tag{6}$$

ii- For $n = 1, 2, , N$, update the codevector

$$C_n^{(i+1)} = \frac{\sum_{Q(X_m)=C_n^{(i)} X_m}}{\sum_{Q(X_m)=C_n^{(i)} 1}} \tag{7}$$

iii- Set i=i+1.

iv Calculate

$$D_{avg}^{(i)} = \frac{1}{Mk} \sum_{m=1}^{M} \left\| X_m - Q\left(X_m\right) \right\| 2 \tag{8}$$

v- If

$$\left(D_{avg}^{(i-1)} - D_{avg}^{(i)}\right) / \left(D_{avg}^{(i-1)} > \varepsilon\right), \tag{9}$$

go back to Step (i).

vi- Set

$$D_{avg}^{*} = D_{avg}^{(i)} \tag{10}$$

For, set

$$D_{avg}^{*} = D_{avg}^{(i)}, \tag{11}$$

as the final codevectors.

5. Repeat Steps 3 and 4 until the desired number of codevectors, hence, clusters is obtained.

Appendix B

Table 1 lists the standardised dataset used to develop the ICT Index featured as an empirical case study in Section 6.3.

UKCI Final Tables Results:-

The Tables: 2, 3, 4, 5, 6, 7, 8 and 9, lists the *UKCI*, EKCI and the TKCI final scores and ranks for 73 economies for the year 2011-2012. The full results for the previous years 2009 and 2010, in addition to the *UKCI* 8 baskets and 2 themes detailed results for the three available years are also listed.

Table 1: Normalised raw data used to calculate the unified ICT index, year 2011.

	[A]	[B]	[C]	[D]	[E]	[F]	[G]	[H]	[I]	[J]	[K]
Argentina	0.374396135	0.417040359	0.657266811	0.536019536	0.155172414	0.257458564	0.077142857	0.959326788	0.396728016	0.249433107	0.273307791
Australia	0.712560386	0.766816143	0.867678959	0.775335775	0.586206897	0.729281768	0.66	0.992987377	0.777096115	0.825396825	0.836526181
Austria	0.782608696	0.701793722	0.938177874	0.831501832	0.689655172	0.874033149	0.614285714	0.995792426	0.781186094	0.607709751	0.762452107
Belgium	0.879227053	0.735426009	0.878524946	0.814407814	0.603448276	0.689502762	0.502857143	0.991584853	0.72392638	0.600907029	0.656449553
Brazil	0.439613527	0.381165919	0.64208243	0.457875458	0.051724138	0.283977901	0.502857143	0.936886396	0.398773006	0.444444444	0.266922095
Bulgaria	0.471014493	0.504484305	0.687635575	0.598290598	0.206896552	0.510497238	0.337142857	0.952314165	0.63599182	0.367346939	0.402298851
Canada	0.782608696	0.878923767	0.85791757	0.800976801	0.862068966	0.693922652	0.657142857	0.992987377	0.72801636	0.825396825	0.619412516
Chile	0.507246377	0.450672646	0.621475054	0.525030525	0.189655172	0.417679558	0.591428571	0.952314165	0.480572597	0.585034014	0.292464879
China	0.34057971	0.392376682	0.376355748	0.365079365	0.672413793	0.477348066	0.737142857	0.960729313	0.404907975	0.530612245	0.218390805
Colombia	0.347826087	0.30044843	0.569414317	0.371184371	0.103448276	0.333701657	0.477142857	0.93828892	0.36809816	0.587301587	0.215836526
Croatia	0.565217391	0.513452915	0.8329718	0.754578755	0.25862069	0.335911602	0.308571429	0.983169705	0.572597137	0.442176871	0.550446999
Czech Republic	0.642512077	0.560538117	0.828633406	0.684981685	0.396551724	0.449723757	0.494285714	0.976157083	0.615541922	0.367346939	0.512132822
Denmark	0.975845411	0.825112108	0.928416486	0.910866911	0.948275862	0.920441989	0.734285714	0.995792426	0.875255624	0.74829932	0.872286079
Estonia	0.673913043	0.634529148	0.880694143	0.737484737	0.413793103	0.594475138	0.688571429	0.981767181	0.744376278	0.725623583	0.519795658
Finland	0.867149758	0.829596413	0.965292842	0.822954823	0.75862069	0.782320442	0.788571429	0.995792426	0.942740286	0.591836735	0.905491699
France	0.838164251	0.737668161	0.85032538	0.84004884	0.655172414	0.689502762	0.602857143	0.990182328	0.705521472	0.698412698	0.730523627
Germany	0.833333333	0.782511211	0.959869848	0.920634921	0.793103448	0.874033149	0.648571429	0.994389902	0.779141104	0.641723356	0.724137931
Greece	0.495169082	0.529147982	0.662689805	0.671550672	0.448275862	0.351381215	0.302857143	0.988779804	0.521472393	0.358276644	0.574712644
Hong Kong, China	0.95410628	0.713004484	0.945770065	1	1	0.796685083	0.697142857	0.998597475	0.828220859	0.795918367	0.82247765
Hungary	0.577294686	0.529147982	0.749457701	0.667887668	0.362068966	0.40441989	0.342857143	0.974754558	0.597137014	0.403628118	0.541507024
Iceland	0.97826087	1	0.911062907	0.981684982	0.844827586	0.910497238	0.668571429	0.997194951	0.858895706	0.458049887	0.837803321
India	0.289855072	0.291479821	0.171366594	0.183150183	0.637931034	0.356906077	0.571428571	0.946704067	0.259713701	0.448979592	0.039591315
Indonesia	0.282608696	0.320627803	0.238611714	0.275946276	0.086206897	0.188950276	0.525714286	0.910238429	0.296523517	0.308390023	0.085568327
Ireland	0.768115942	0.683856502	0.855748373	0.803418803	0.551724138	0.837569061	0.471428571	0.991584853	0.658486708	0.528344671	0.657726692
Italy	0.526570048	0.515695067	0.855748373	0.73992674	0.431034483	0.561325967	0.274285714	0.991584853	0.617586912	0.31292517	0.634738186
Japan	0.700483092	0.67264574	0.840563991	0.765567766	0.568965517	0.815469613	0.602857143	0.988779804	0.791411043	0.678004535	0.901660281
Jordan	0.398550725	0.35426009	0.457700651	0.421245421	0.172413793	0.260773481	0.577142857	0.949509116	0.406952965	0.535147392	0.191570881
Kazakhstan	0.309178744	0.347533632	0.542299349	0.456654457	0.293103448	0.299447514	0.468571429	0.980364656	0.351738241	0.553287982	0.181353768
Korea Rep.	0.765700483	0.760089686	0.838394794	0.896214896	0.775862069	0.667403315	0.682857143	0.991584853	0.887525562	1	1
Lithuania	0.61352657	0.542600897	0.788503254	0.684981685	0.5	0.495027624	0.397142857	0.987377279	0.644171779	0.555555556	0.504469987
Luxembourg	0.927536232	0.724215247	0.992407809	0.968253968	0.879310345	0.321546961	0.811428571	0.997194951	0.918200409	0.489795918	0.922094508
Malaysia	0.514492754	0.432735426	0.682212581	0.467643468	0.706896552	0.575690608	0.771428571	0.978962132	0.552147239	0.739229025	0.399744572
Mexico	0.384057971	0.329596413	0.469631236	0.374847375	0.068965517	0.303867403	0.337142857	0.970546985	0.331288344	0.426303855	0.234993614
Netherlands	0.958937198	0.867713004	0.990238612	0.905982906	0.827586207	0.727071823	0.608571429	0.994389902	0.852760736	0.680272109	0.812260536
New Zealand	0.710144928	0.717488789	0.865509761	0.813186813	0.534482759	0.811049724	0.648571429	0.981767181	0.795501022	0.718820862	0.808429119
Norway	0.946859903	0.843049327	0.890455531	0.855921856	0.810344828	0.523756906	0.657142857	0.997194951	0.838445808	0.712018141	0.840357599
Peru	0.359903382	0.286995516	0.527114967	0.335775336	0.034482759	0.467403315	0.322857143	0.870967742	0.325153374	0.346938776	0.196679438
Philippines	0.316425121	0.266816143	0.293926247	0.277167277	0.465517241	0.166850829	0.254285714	0.875175316	0.308793456	0.308390023	0.187739464
Poland	0.487922705	0.47309417	0.69197397	0.686202686	0.327586207	0.370165746	0.317142857	0.978962132	0.54601227	0.287981859	0.487867178
Portugal	0.760869565	0.535874439	0.768980477	0.765567766	0.620689655	0.561325967	0.797142857	0.98457223	0.693251534	0.555555556	0.660280971
Qatar	0.623188406	0.529147982	0.686550976	0.759462759	0.344827586	0.212154696	0.971428571	0.966339411	0.685071575	0.514739229	0.476372925
Romania	0.38647343	0.459641256	0.636659436	0.565323565	0.275862069	0.366850829	0.262857143	0.977559607	0.503067485	0.319727891	0.406130268
Russian Federation	0.362318841	0.47309417	0.74186551	0.672771673	0.310344828	0.256353591	0.382857143	0.988779804	0.480572597	0.274376417	0.332056194
Singapore	0.903381643	0.780269058	0.917570499	0.887667888	0.965517241	0.654143646	1	0.997194951	0.852760736	0.863945578	0.767560664
Slovak Republic	0.574879227	0.448430493	0.798264642	0.645909646	0.137931034	0.359116022	0.197142857	0.974754558	0.629856851	0.22675737	0.56449553
Slovenia	0.628019324	0.621076233	0.811279826	0.774114774	0.379310345	0.318232044	0.485714286	0.987377279	0.678936605	0.541950113	0.607918263
South Africa	0.347826087	0.32735426	0.353579176	0.278388278	0.017241379	0.650828729	0.354285714	0.929873773	0.26993865	0.317460317	0.130268199
Spain	0.673913043	0.594170404	0.803687636	0.746031746	0.482758621	0.727071823	0.351428571	0.985974755	0.658486708	0.70521542	0.680715198
Sweden	0.997584541	0.968609865	0.994577007	0.94017094	0.896551724	0.571270718	0.811428571	0.995792426	1	0.696145125	0.961685824
Switzerland	1	0.899103139	0.963123644	0.956043956	0.74137931	0.8	0.671428571	0.994389902	0.844580777	0.489795918	0.810983397
Thailand	0.316425121	0.266816143	0.567245119	0.335775336	0.120689655	0.428729282	0.454285714	0.949509116	0.357873211	0.35600907	0.131545338
Turkey	0.43236715	0.390134529	0.453362256	0.500610501	0.224137931	0.260773481	0.468571429	0.952314165	0.421267894	0.385487528	0.311621967
Ukraine	0.316425121	0.441704036	0.503253796	0.478632479	0.24137931	0.390055249	0.225714286	0.967741935	0.316973415	0.319727891	0.169859515
United Arab Emirates	0.657004831	0.609865471	0.928416486	0.719169719	0.517241379	0.499447514	0.882857143	0.998597475	0.748466258	0.507936508	0.651340996
United Kingdom	0.946859903	0.84529148	0.990238612	0.914529915	0.724137931	1	0.597142857	0.994389902	0.81595092	0.746031746	0.819923372
United States	0.741545894	0.876681614	0.888286334	0.777777778	0.982758621	0.910497238	0.725714286	0.995792426	0.760736196	0.854875283	0.749680715
Venezuela	0.289855072	0.352017937	0.584598698	0.361416361	0	0.009944751	0.12	0.981767181	0.316973415	0.19047619	0.275862069

Table 2: UKCI final aggregated scores and ranks for 73 economies, year 2010-2011.

Countries	UKCI _Final 2010-2011		Explicit Knowledge 2010-2011		Tacit Knowledge 2010-2011	
	Score	Rank	Score	Rank	Score	Rank
Sweden	**0.876**	1	0.866	1	0.897	2
Switzerland	**0.871**	2	0.861	2	0.909	1
Denmark	**0.834**	3	0.844	3	0.817	7
Singapore	**0.825**	4	0.837	5	0.785	9
Finland	**0.812**	5	0.816	7	0.853	4
Canada	**0.808**	6	0.838	4	0.776	10
United States	**0.806**	7	0.792	11	0.857	3
Norway	**0.802**	8	0.826	6	0.763	12
Netherlands	**0.801**	9	0.801	8	0.801	8
Germany	**0.781**	10	0.792	10	0.835	6
Australia	**0.755**	11	0.749	14	0.697	18
Austria	**0.749**	12	0.757	13	0.739	15
Taiwan China	**0.748**	13	0.717	18	0.765	11
Hong Kong	**0.731**	14	0.741	15	0.589	27
Japan	**0.723**	15	0.700	20	0.836	5
United Kingdom	**0.720**	16	0.731	16	0.739	16
Luxembourg	**0.714**	17	0.777	12	0.544	32
France	**0.697**	18	0.728	17	0.719	17
New Zealand	**0.696**	19	0.714	19	0.662	24
Iceland	**0.695**	20	0.796	9	0.695	19
Belgium	**0.691**	21	0.676	21	0.749	14
Israel	**0.668**	22	0.625	24	0.750	13
Ireland	**0.660**	23	0.631	22	0.688	20
Korea Rep.	**0.642**	24	0.627	23	0.670	22
Malaysia	**0.609**	25	0.560	28	0.608	26
Qatar	**0.602**	26	0.502	31	0.680	21
United Arab Emirates	**0.595**	27	0.599	25	0.665	23
Czech Republic	**0.559**	28	0.561	27	0.617	25
Estonia	**0.538**	29	0.561	26	0.497	36
Spain	**0.531**	30	0.535	29	0.547	31
Chile	**0.527**	31	0.487	34	0.502	34
China	**0.508**	32	0.475	36	0.575	29
Saudi Arabia	**0.503**	33	0.407	43	0.550	30
Slovenia	**0.499**	34	0.501	32	0.579	28
Bahrain	**0.487**	35	0.491	33	0.375	51
Portugal	**0.482**	36	0.515	30	0.464	43
Italy	**0.470**	37	0.439	38	0.530	33
Lithuania	**0.468**	38	0.483	35	0.485	37
Oman	**0.460**	39	0.405	44	0.454	44
Poland	**0.457**	40	0.410	42	0.468	42
Hungary	**0.446**	41	0.438	39	0.473	38
Slovak Republic	**0.441**	42	0.421	40	0.472	39
Kuwait	**0.440**	43	0.441	37	0.359	59
Tunisia	**0.423**	44	0.369	45	0.437	46
South Africa	**0.421**	45	0.353	47	0.472	40
Thailand	**0.413**	46	0.348	48	0.402	48
India	**0.413**	47	0.287	57	0.501	35
Brazil	**0.411**	48	0.306	53	0.471	41
Greece	**0.405**	49	0.419	41	0.379	50
Turkey	**0.365**	50	0.326	51	0.368	56
Kazakhstan	**0.361**	51	0.339	50	0.327	62
Russian Federation	**0.354**	52	0.341	49	0.440	45
Lebanon	**0.352**	53	0.293	56	0.372	53
Jordan	**0.349**	54	0.299	55	0.391	49
Croatia	**0.348**	55	0.355	46	0.363	57
Indonesia	**0.339**	56	0.240	65	0.419	47
Egypt	**0.337**	57	0.276	59	0.359	58
Colombia	**0.333**	58	0.266	61	0.369	55
Romania	**0.333**	59	0.305	54	0.357	60
Bulgaria	**0.321**	60	0.324	52	0.314	63
Iran	**0.314**	61	0.276	60	0.266	66
Morocco	**0.311**	62	0.280	58	0.250	68
Argentina	**0.308**	63	0.251	63	0.373	52
Mexico	**0.299**	64	0.251	64	0.336	61
Philippines	**0.288**	65	0.217	66	0.305	64
Ukraine	**0.288**	66	0.261	62	0.372	54
Peru	**0.274**	67	0.205	67	0.283	65
Algeria	**0.242**	68	0.198	68	0.233	70
Syria	**0.221**	69	0.197	69	0.183	71
Libya	**0.204**	70	0.184	71	0.169	72
Mauritania	**0.203**	71	0.186	70	0.242	69
Venezuela	**0.188**	72	0.130	73	0.258	67
Yemen	**0.156**	73	0.159	72	0.127	73

Table 3: UKCI final aggregated scores and ranks for 73 economies, year 2009-2010.

Countries	UKCI _Final 2009-2010		Explicit Knowledge 2009-2010		Tacit Knowledge 2009-2010	
	Score	Rank	Score	Rank	Score	Rank
Switzerland	**0.914**	1	0.896	2	0.955	1
United States	**0.884**	2	0.872	3	0.944	2
Denmark	**0.876**	3	0.925	1	0.869	6
Sweden	**0.873**	4	0.835	5	0.905	3
Singapore	**0.828**	5	0.846	4	0.823	9
Germany	**0.824**	6	0.728	16	0.891	4
Finland	**0.815**	7	0.777	7	0.837	8
Japan	**0.799**	8	0.726	17	0.890	5
Netherlands	**0.795**	9	0.744	13	0.805	13
Canada	**0.792**	10	0.823	6	0.764	14
United Kingdom	**0.782**	11	0.750	12	0.821	10
Austria	**0.775**	12	0.759	10	0.810	12
Norway	**0.755**	13	0.738	15	0.737	17
Taiwan China	**0.754**	14	0.704	21	0.818	11
Korea Rep.	**0.753**	15	0.711	20	0.850	7
Hong Kong	**0.738**	16	0.638	23	0.709	20
Belgium	**0.730**	17	0.776	8	0.760	15
Iceland	**0.722**	18	0.716	19	0.691	21
France	**0.722**	19	0.689	22	0.751	16
Australia	**0.721**	20	0.762	9	0.659	24
Israel	**0.706**	21	0.741	14	0.726	19
Ireland	**0.698**	22	0.724	18	0.726	18
Luxembourg	**0.696**	23	0.587	30	0.662	23
New Zealand	**0.681**	24	0.754	11	0.619	25
Malaysia	**0.618**	25	0.621	24	0.665	22
Czech Republic	**0.584**	26	0.610	27	0.594	26
Qatar	**0.581**	27	0.584	31	0.576	27
United Arab Emirates	**0.577**	28	0.578	33	0.569	29
Estonia	**0.564**	29	0.616	25	0.523	34
Slovenia	**0.560**	30	0.598	29	0.548	30
Spain	**0.538**	31	0.528	39	0.536	33
China	**0.521**	32	0.514	41	0.536	32
Italy	**0.519**	33	0.480	47	0.574	28
Portugal	**0.502**	34	0.545	36	0.476	40
Chile	**0.497**	35	0.534	38	0.472	41
Lithuania	**0.497**	36	0.570	34	0.451	46
Kuwait	**0.496**	37	0.599	28	0.477	39
Saudi Arabia	**0.490**	38	0.513	42	0.466	42
Slovak Republic	**0.488**	39	0.515	40	0.463	43
Hungary	**0.475**	40	0.559	35	0.448	47
Bahrain	**0.471**	41	0.470	52	0.357	53
India	**0.459**	42	0.487	46	0.547	31
Tunisia	**0.458**	43	0.579	32	0.481	38
Thailand	**0.447**	44	0.489	45	0.454	45
Oman	**0.445**	45	0.475	49	0.428	48
Poland	**0.441**	46	0.535	37	0.420	50
Brazil	**0.439**	47	0.473	51	0.493	35
South Africa	**0.439**	48	0.475	50	0.490	36
Jordan	**0.424**	49	0.466	53	0.456	44
Indonesia	**0.412**	50	0.506	44	0.482	37
Turkey	**0.397**	51	0.403	59	0.425	49
Russian Federation	**0.392**	52	0.508	43	0.327	60
Greece	**0.389**	53	0.432	55	0.336	57
Lebanon	**0.384**	54	0.611	26	0.349	54
Croatia	**0.372**	55	0.409	57	0.344	55
Kazakhstan	**0.361**	56	0.456	54	0.309	62
Romania	**0.355**	57	0.403	60	0.334	59
Bulgaria	**0.347**	58	0.410	56	0.282	65
Mexico	**0.343**	59	0.408	58	0.367	52
Colombia	**0.336**	60	0.400	61	0.336	58
Ukraine	**0.335**	61	0.367	64	0.339	56
Iran	**0.328**	62	0.479	48	0.274	66
Philippines	**0.327**	63	0.342	66	0.382	51
Egypt	**0.312**	64	0.385	63	0.297	64
Morocco	**0.308**	65	0.328	68	0.309	63
Argentina	**0.298**	66	0.398	62	0.318	61
Peru	**0.288**	67	0.343	65	0.267	67
Syria	**0.253**	68	0.336	67	0.238	68
Algeria	**0.224**	69	0.302	69	0.131	72
Venezuela	**0.204**	70	0.278	71	0.171	69
Libya	**0.188**	71	0.279	70	0.141	71
Mauritania	**0.165**	72	0.165	72	0.149	70
Yemen	**0.130**	73	0.132	73	0.107	73

Appendix B

Table 4: Explicit Macro-Knowledge Competitiveness Indicator (EKCI) and the theme detailed baskets scores, year 2011-2012.

Countries	Explicit Macro-Knowledge Score	Explicit Macro-Knowledge Rank	Economy Score	Education Score	Public Admin Score	Infrastructure Score	Labor Score
Switzerland	0.911	1	0.750	0.817	0.919	0.873	0.970
Singapore	0.884	2	0.847	0.805	0.928	0.813	0.951
Denmark	0.878	3	0.745	0.905	0.946	0.836	0.853
Sweden	0.872	4	0.818	0.879	0.925	0.877	0.781
Finland	0.853	5	0.690	0.962	0.963	0.792	0.799
Canada	0.826	6	0.793	0.865	0.866	0.816	0.783
Netherlands	0.818	7	0.708	0.865	0.849	0.833	0.740
Hong Kong	0.816	8	0.898	0.757	0.829	0.823	0.781
Iceland	0.816	9	0.386	0.923	0.769	0.872	0.780
Luxembourg	0.805	10	0.719	0.680	0.849	0.766	0.813
Norway	0.796	11	0.723	0.880	0.864	0.786	0.709
United States	0.784	12	0.808	0.839	0.753	0.782	0.836
United Kingdom	0.783	13	0.595	0.798	0.788	0.775	0.793
Australia	0.776	14	0.776	0.883	0.850	0.724	0.760
Germany	0.774	15	0.711	0.832	0.803	0.837	0.611
Austria	0.763	16	0.620	0.804	0.842	0.744	0.677
New Zealand	0.751	17	0.580	0.913	0.883	0.666	0.705
Belgium	0.744	18	0.608	0.888	0.758	0.740	0.730
Japan	0.732	19	0.569	0.793	0.720	0.705	0.802
France	0.721	20	0.441	0.803	0.724	0.758	0.646
Ireland	0.705	21	0.590	0.849	0.722	0.673	0.740
Taiwan China	0.685	22	0.814	0.864	0.640	0.744	0.644
Israel	0.680	23	0.645	0.763	0.650	0.642	0.799
Korea Rep.	0.649	24	0.600	0.867	0.606	0.744	0.535
Estonia	0.623	25	0.453	0.808	0.665	0.557	0.683
United Arab Emirates	0.613	26	0.413	0.684	0.579	0.605	0.681
Czech Republic	0.609	27	0.440	0.750	0.589	0.576	0.701
Spain	0.584	28	0.435	0.748	0.576	0.645	0.483
Malaysia	0.582	29	0.727	0.652	0.575	0.532	0.690
Bahrain	0.564	30	0.556	0.758	0.593	0.583	0.485
Qatar	0.563	31	0.708	0.613	0.603	0.547	0.530
Slovenia	0.558	32	0.249	0.828	0.556	0.561	0.555
Portugal	0.558	33	0.349	0.746	0.611	0.607	0.382
Chile	0.550	34	0.627	0.625	0.666	0.447	0.565
Lithuania	0.510	35	0.315	0.787	0.516	0.496	0.528
Saudi Arabia	0.499	36	0.637	0.677	0.489	0.522	0.471
Slovak Republic	0.494	37	0.349	0.677	0.530	0.460	0.503
Hungary	0.494	38	0.295	0.731	0.499	0.503	0.471
Italy	0.494	39	0.357	0.769	0.443	0.527	0.512
Poland	0.483	40	0.399	0.761	0.500	0.434	0.551
Oman	0.475	41	0.555	0.573	0.587	0.413	0.417
China	0.443	42	0.598	0.627	0.330	0.434	0.636
Croatia	0.436	43	0.205	0.666	0.421	0.469	0.396
Kuwait	0.433	44	0.579	0.558	0.495	0.437	0.328
Greece	0.431	45	0.195	0.780	0.424	0.496	0.317
Thailand	0.385	46	0.632	0.534	0.336	0.330	0.566
Russian Federation	0.374	47	0.257	0.692	0.173	0.424	0.591
Bulgaria	0.373	48	0.237	0.622	0.333	0.363	0.455
Tunisia	0.372	49	0.445	0.661	0.496	0.352	0.219
Turkey	0.365	50	0.489	0.521	0.372	0.386	0.314
Lebanon	0.356	51	0.412	0.655	0.395	0.249	0.502
South Africa	0.355	52	0.472	0.424	0.433	0.249	0.437
Brazil	0.353	53	0.466	0.562	0.312	0.333	0.458
Kazakhstan	0.343	54	0.395	0.631	0.302	0.301	0.486
Romania	0.339	55	0.317	0.678	0.308	0.354	0.361
Jordan	0.330	56	0.270	0.636	0.397	0.308	0.267
Egypt	0.317	57	0.435	0.456	0.378	0.300	0.256
Morocco	0.315	58	0.449	0.372	0.435	0.288	0.177
Mexico	0.314	59	0.391	0.551	0.312	0.297	0.348
Colombia	0.302	60	0.388	0.558	0.288	0.274	0.376
Peru	0.294	61	0.424	0.497	0.235	0.230	0.507
Iran	0.292	62	0.446	0.543	0.348	0.315	0.159
India	0.290	63	0.583	0.390	0.351	0.233	0.302
Argentina	0.280	64	0.241	0.656	0.210	0.316	0.320
Ukraine	0.280	65	0.222	0.765	0.122	0.307	0.475
Philippines	0.272	66	0.422	0.472	0.269	0.203	0.409
Syria	0.267	67	0.279	0.471	0.313	0.229	0.267
Algeria	0.253	68	0.346	0.521	0.316	0.232	0.194
Indonesia	0.236	69	0.423	0.535	0.283	0.235	0.168
Mauritania	0.220	70	0.175	0.096	0.304	0.090	0.338
Yemen	0.174	71	0.222	0.186	0.253	0.069	0.252
Libya	0.169	72	0.254	0.420	0.200	0.203	0.055
Venezuela	0.143	73	0.135	0.626	0.026	0.180	0.254

Table 5: Explicit Macro-Knowledge Competitiveness Indicator (EKCI) and the theme detailed baskets scores, year 2010-2011.

Countries	Explicit Macro-Knowledge Score	Explicit Macro-Knowledge Rank	Economy Score	Education Score	Public Admin Score	Infrastructure Score	Labor Score
Sweden	0.866	1	0.786	0.914	0.911	0.868	0.747
Switzerland	0.861	2	0.791	0.848	0.893	0.840	0.867
Denmark	0.844	3	0.744	0.926	0.928	0.820	0.739
Canada	0.838	4	0.780	0.898	0.851	0.837	0.812
Singapore	0.837	5	0.823	0.790	0.946	0.783	0.784
Norway	0.826	6	0.764	0.884	0.857	0.815	0.795
Finland	0.816	7	0.649	0.990	0.909	0.775	0.756
Netherlands	0.801	8	0.713	0.883	0.834	0.792	0.752
Iceland	0.796	9	0.337	0.929	0.708	0.861	0.753
Germany	0.792	10	0.623	0.807	0.771	0.822	0.727
United States	0.792	11	0.783	0.857	0.720	0.824	0.843
Luxembourg	0.777	12	0.756	0.683	0.848	0.698	0.917
Austria	0.757	13	0.698	0.820	0.796	0.751	0.683
Australia	0.749	14	0.822	0.894	0.854	0.681	0.766
Hong Kong	0.741	15	0.851	0.730	0.841	0.724	0.564
United Kingdom	0.731	16	0.579	0.819	0.730	0.736	0.713
France	0.728	17	0.527	0.831	0.696	0.765	0.661
Taiwan China	0.717	18	0.811	0.851	0.632	0.781	0.668
New Zealand	0.714	19	0.585	0.907	0.882	0.603	0.747
Japan	0.700	20	0.569	0.818	0.652	0.720	0.741
Belgium	0.676	21	0.549	0.903	0.691	0.666	0.680
Ireland	0.631	22	0.578	0.840	0.719	0.568	0.670
Korea Rep.	0.627	23	0.585	0.860	0.521	0.711	0.550
Israel	0.625	24	0.639	0.685	0.571	0.623	0.763
United Arab Emirates	0.599	25	0.388	0.661	0.591	0.549	0.820
Estonia	0.561	26	0.407	0.808	0.618	0.526	0.561
Czech Republic	0.561	27	0.397	0.767	0.540	0.533	0.721
Malaysia	0.560	28	0.783	0.615	0.582	0.525	0.649
Spain	0.535	29	0.410	0.745	0.520	0.546	0.533
Portugal	0.515	30	0.314	0.704	0.555	0.522	0.389
Qatar	0.502	31	0.659	0.667	0.625	0.429	0.492
Slovenia	0.501	32	0.217	0.840	0.500	0.474	0.609
Bahrain	0.491	33	0.555	0.670	0.550	0.544	0.135
Chile	0.487	34	0.623	0.635	0.657	0.340	0.656
Lithuania	0.483	35	0.285	0.780	0.472	0.451	0.635
China	0.475	36	0.587	0.588	0.378	0.461	0.767
Kuwait	0.441	37	0.596	0.540	0.488	0.417	0.422
Italy	0.439	38	0.356	0.761	0.388	0.448	0.531
Hungary	0.438	39	0.320	0.708	0.430	0.413	0.559
Slovak Republic	0.421	40	0.329	0.667	0.442	0.344	0.679
Greece	0.419	41	0.333	0.777	0.381	0.438	0.438
Poland	0.410	42	0.459	0.770	0.453	0.341	0.581
Saudi Arabia	0.407	43	0.630	0.590	0.456	0.448	0.119
Oman	0.405	44	0.578	0.492	0.584	0.344	0.209
Tunisia	0.369	45	0.463	0.644	0.518	0.332	0.151
Croatia	0.355	46	0.164	0.654	0.349	0.340	0.428
South Africa	0.353	47	0.515	0.371	0.483	0.242	0.479
Thailand	0.348	48	0.618	0.578	0.351	0.322	0.444
Russian Federation	0.341	49	0.226	0.705	0.175	0.359	0.671
Kazakhstan	0.339	50	0.459	0.598	0.340	0.286	0.548
Turkey	0.326	51	0.426	0.513	0.328	0.317	0.355
Bulgaria	0.324	52	0.215	0.638	0.314	0.288	0.489
Brazil	0.306	53	0.561	0.560	0.294	0.273	0.470
Romania	0.305	54	0.263	0.674	0.298	0.254	0.523
Jordan	0.299	55	0.322	0.613	0.380	0.216	0.432
Lebanon	0.293	56	0.412	0.636	0.366	0.206	0.462
India	0.287	57	0.649	0.390	0.333	0.196	0.539
Morocco	0.280	58	0.458	0.303	0.413	0.236	0.131
Egypt	0.276	59	0.441	0.447	0.380	0.268	0.058
Iran	0.276	60	0.474	0.523	0.324	0.250	0.262
Colombia	0.266	61	0.398	0.553	0.258	0.213	0.493
Ukraine	0.261	62	0.171	0.737	0.122	0.271	0.559
Argentina	0.251	63	0.265	0.657	0.200	0.236	0.438
Mexico	0.251	64	0.309	0.512	0.259	0.217	0.366
Indonesia	0.240	65	0.441	0.529	0.293	0.156	0.443
Philippines	0.217	66	0.405	0.519	0.248	0.159	0.375
Peru	0.205	67	0.380	0.519	0.235	0.125	0.449
Algeria	0.198	68	0.342	0.452	0.280	0.176	0.085
Syria	0.197	69	0.288	0.432	0.288	0.180	0.047
Mauritania	0.186	70	0.186	0.059	0.286	0.107	0.259
Libya	0.184	71	0.296	0.350	0.250	0.178	0.043
Yemen	0.159	72	0.221	0.138	0.209	0.129	0.159
Venezuela	0.130	73	0.156	0.577	0.031	0.112	0.445

Table 6: Explicit Macro-Knowledge Competitiveness Indicator (EKCI) and the theme detailed baskets scores, year 2009-2010.

Countries	Explicit Macro-Knowledge Score	Explicit Macro-Knowledge Rank	Economy Score	Education Score	Public Admin Score	Infrastructure Score	Labor Score
Denmark	0.925	1	0.763	0.950	0.943	0.862	0.804
Switzerland	0.896	2	0.787	0.897	0.862	0.879	0.904
United States	0.872	3	0.787	0.881	0.658	0.887	0.898
Singapore	0.846	4	0.751	0.834	0.913	0.779	0.886
Sweden	0.835	5	0.733	0.860	0.869	0.907	0.709
Canada	0.823	6	0.699	0.824	0.831	0.852	0.819
Finland	0.777	7	0.754	0.782	0.936	0.828	0.704
Belgium	0.776	8	0.569	0.815	0.666	0.718	0.634
Australia	0.762	9	0.777	0.755	0.845	0.702	0.767
Austria	0.759	10	0.692	0.782	0.796	0.733	0.641
New Zealand	0.754	11	0.669	0.748	0.885	0.623	0.742
United Kingdom	0.750	12	0.628	0.758	0.678	0.750	0.737
Netherlands	0.744	13	0.712	0.750	0.812	0.816	0.694
Israel	0.741	14	0.623	0.762	0.607	0.621	0.692
Norway	0.738	15	0.649	0.731	0.817	0.813	0.747
Germany	0.728	16	0.678	0.752	0.762	0.873	0.604
Japan	0.726	17	0.649	0.730	0.642	0.772	0.736
Ireland	0.724	18	0.586	0.738	0.725	0.596	0.659
Iceland	0.716	19	0.444	0.702	0.706	0.848	0.795
Korea Rep.	0.711	20	0.586	0.773	0.479	0.752	0.493
Taiwan China	0.704	21	0.548	0.721	0.605	0.754	0.661
France	0.689	22	0.599	0.713	0.673	0.770	0.578
Hong Kong	0.638	23	0.834	0.606	0.855	0.680	0.711
Malaysia	0.621	24	0.647	0.626	0.529	0.546	0.629
Estonia	0.616	25	0.410	0.618	0.658	0.510	0.600
Lebanon	0.611	26	0.437	0.667	0.334	0.216	0.441
Czech Republic	0.610	27	0.511	0.604	0.456	0.542	0.694
Kuwait	0.599	28	0.561	0.635	0.475	0.414	0.467
Slovenia	0.598	29	0.476	0.609	0.512	0.478	0.575
Luxembourg	0.587	30	0.701	0.534	0.823	0.676	0.759
Qatar	0.584	31	0.603	0.576	0.658	0.475	0.594
Tunisia	0.579	32	0.472	0.630	0.532	0.337	0.348
United Arab Emirates	0.578	33	0.448	0.538	0.607	0.452	0.760
Lithuania	0.570	34	0.410	0.569	0.477	0.466	0.607
Hungary	0.559	35	0.396	0.576	0.413	0.404	0.531
Portugal	0.545	36	0.435	0.579	0.528	0.492	0.388
Poland	0.535	37	0.383	0.541	0.392	0.294	0.560
Chile	0.534	38	0.610	0.506	0.641	0.332	0.627
Spain	0.528	39	0.488	0.543	0.483	0.506	0.471
Slovak Republic	0.515	40	0.587	0.492	0.414	0.300	0.656
China	0.514	41	0.547	0.485	0.388	0.564	0.693
Saudi Arabia	0.513	42	0.583	0.523	0.534	0.441	0.456
Russian Federation	0.508	43	0.308	0.502	0.211	0.383	0.636
Indonesia	0.506	44	0.427	0.534	0.270	0.233	0.454
Thailand	0.489	45	0.513	0.484	0.410	0.360	0.535
India	0.487	46	0.642	0.495	0.339	0.239	0.493
Italy	0.480	47	0.416	0.506	0.338	0.416	0.406
Iran	0.479	48	0.412	0.533	0.320	0.264	0.276
Oman	0.475	49	0.558	0.454	0.606	0.304	0.533
South Africa	0.475	50	0.583	0.481	0.453	0.206	0.447
Brazil	0.473	51	0.482	0.488	0.277	0.226	0.465
Bahrain	0.470	52	0.606	0.441	0.584	0.510	0.568
Jordan	0.466	53	0.411	0.482	0.459	0.295	0.394
Kazakhstan	0.456	54	0.379	0.433	0.331	0.329	0.609
Greece	0.432	55	0.339	0.452	0.337	0.414	0.373
Bulgaria	0.410	56	0.337	0.396	0.321	0.283	0.508
Croatia	0.409	57	0.270	0.414	0.315	0.273	0.422
Mexico	0.408	58	0.406	0.433	0.268	0.139	0.335
Turkey	0.403	59	0.484	0.430	0.302	0.266	0.305
Romania	0.403	60	0.341	0.391	0.311	0.199	0.490
Colombia	0.400	61	0.365	0.395	0.264	0.184	0.472
Argentina	0.398	62	0.252	0.429	0.150	0.186	0.337
Egypt	0.385	63	0.401	0.418	0.372	0.269	0.233
Ukraine	0.367	64	0.248	0.345	0.137	0.252	0.556
Peru	0.343	65	0.413	0.326	0.275	0.118	0.446
Philippines	0.342	66	0.387	0.348	0.226	0.151	0.349
Syria	0.336	67	0.261	0.372	0.287	0.193	0.183
Morocco	0.328	68	0.419	0.358	0.392	0.245	0.158
Algeria	0.302	69	0.263	0.322	0.269	0.212	0.220
Libya	0.279	70	0.275	0.328	0.216	0.205	0.061
Venezuela	0.278	71	0.244	0.302	0.077	0.112	0.232
Mauritania	0.165	72	0.201	0.123	0.290	0.077	0.325
Yemen	0.132	73	0.218	0.117	0.221	0.146	0.170

Table 7: Tacit Macro-Knowledge Competitiveness Indicator (TKCI) and the theme detailed baskets scores, year 2011-2012.

Countries	Tacit Macro-Knowledge Score	Tacit Macro-Knowledge Rank	ICT and E-Services Score	Intellectual Capital Score	Innovation Score
Sweden	0.898	1	0.933	0.945	0.892
Switzerland	0.869	2	0.903	0.959	0.858
Singapore	0.821	3	0.828	0.797	0.823
Taiwan China	0.817	4	0.785	0.761	0.824
Finland	0.811	5	0.873	0.853	0.805
United States	0.808	6	0.775	0.806	0.809
Netherlands	0.807	7	0.845	0.837	0.803
Germany	0.799	8	0.831	0.854	0.792
Denmark	0.762	9	0.904	0.842	0.751
Japan	0.753	10	0.757	0.868	0.739
Israel	0.751	11	0.745	0.781	0.748
United Kingdom	0.751	12	0.870	0.790	0.745
Canada	0.751	13	0.762	0.716	0.755
Hong Kong	0.732	14	0.910	0.703	0.734
Ireland	0.726	15	0.720	0.715	0.727
Luxembourg	0.705	16	0.914	0.653	0.710
Korea Rep.	0.704	17	0.864	0.697	0.703
Norway	0.688	18	0.835	0.735	0.681
France	0.677	19	0.773	0.761	0.666
Austria	0.673	20	0.790	0.795	0.657
Belgium	0.670	21	0.773	0.787	0.655
Malaysia	0.592	22	0.481	0.724	0.577
Iceland	0.592	23	0.913	0.699	0.576
Australia	0.587	24	0.760	0.672	0.576
Hungary	0.581	25	0.596	0.439	0.598
China	0.574	26	0.359	0.600	0.573
Czech Republic	0.569	27	0.623	0.620	0.562
Estonia	0.540	28	0.702	0.566	0.535
New Zealand	0.540	29	0.778	0.658	0.524
Italy	0.526	30	0.625	0.627	0.513
Spain	0.514	31	0.662	0.595	0.503
Slovenia	0.494	32	0.679	0.565	0.484
Brazil	0.487	33	0.403	0.587	0.476
Portugal	0.482	34	0.711	0.523	0.475
Thailand	0.453	35	0.291	0.522	0.446
India	0.452	36	0.218	0.505	0.448
Poland	0.435	37	0.538	0.476	0.430
Russian Federation	0.406	38	0.491	0.344	0.413
United Arab Emirates	0.399	39	0.683	0.734	0.357
Qatar	0.389	40	0.663	0.840	0.333
Lebanon	0.389	41	0.292	0.522	0.374
Argentina	0.387	42	0.414	0.453	0.379
Lithuania	0.385	43	0.601	0.489	0.371
Romania	0.384	44	0.477	0.340	0.388
Oman	0.382	45	0.480	0.504	0.367
Chile	0.381	46	0.474	0.583	0.356
Turkey	0.376	47	0.422	0.445	0.367
Slovak Republic	0.375	48	0.598	0.443	0.365
Greece	0.374	49	0.554	0.360	0.374
Croatia	0.373	50	0.614	0.389	0.369
Saudi Arabia	0.367	51	0.591	0.687	0.326
South Africa	0.365	52	0.279	0.534	0.346
Kuwait	0.363	53	0.395	0.360	0.363
Tunisia	0.350	54	0.358	0.507	0.331
Ukraine	0.347	55	0.333	0.334	0.349
Mexico	0.346	56	0.334	0.424	0.337
Colombia	0.333	57	0.343	0.468	0.316
Jordan	0.330	58	0.377	0.401	0.321
Bulgaria	0.320	59	0.549	0.358	0.313
Philippines	0.312	60	0.287	0.429	0.299
Indonesia	0.302	61	0.247	0.453	0.284
Bahrain	0.291	62	0.651	0.502	0.263
Iran	0.279	63	0.276	0.285	0.278
Egypt	0.270	64	0.308	0.374	0.258
Venezuela	0.266	65	0.291	0.307	0.261
Morocco	0.250	66	0.345	0.333	0.239
Mauritania	0.240	67	0.101	0.167	0.250
Kazakhstan	0.226	68	0.358	0.314	0.215
Peru	0.216	69	0.303	0.427	0.190
Syria	0.180	70	0.230	0.270	0.168
Libya	0.148	71	0.231	0.072	0.156
Algeria	0.085	72	0.215	0.195	0.071
Yemen	0.085	73	0.047	0.172	0.075

Table 8: Tacit Macro-Knowledge Competitiveness Indicator (TKCI) and the theme detailed baskets scores, year 2010-2011.

Countries	Tacit Macro-Knowledge Score	Tacit Macro-Knowledge Rank	ICT and E-Services Score	Intellectual Capital Score	Innovation Score
Switzerland	0.909	1	0.909	0.929	0.889
Sweden	0.897	2	0.973	0.943	0.851
United States	0.857	3	0.776	0.833	0.882
Finland	0.853	4	0.792	0.866	0.841
Japan	0.836	5	0.709	0.851	0.822
Germany	0.835	6	0.841	0.855	0.815
Denmark	0.817	7	0.897	0.870	0.765
Netherlands	0.801	8	0.939	0.827	0.775
Singapore	0.785	9	0.840	0.773	0.796
Canada	0.776	10	0.796	0.752	0.800
Taiwan China	0.765	11	0.788	0.751	0.778
Norway	0.763	12	0.848	0.773	0.752
Israel	0.750	13	0.718	0.684	0.816
Belgium	0.749	14	0.742	0.775	0.723
Austria	0.739	15	0.774	0.765	0.715
United Kingdom	0.739	16	0.873	0.731	0.747
France	0.719	17	0.765	0.741	0.697
Australia	0.697	18	0.741	0.664	0.731
Iceland	0.695	19	0.912	0.711	0.679
Ireland	0.688	20	0.730	0.691	0.686
Qatar	0.680	21	0.569	0.670	0.690
Korea Rep.	0.670	22	0.799	0.649	0.691
United Arab Emirates	0.665	23	0.716	0.667	0.664
New Zealand	0.662	24	0.723	0.657	0.667
Czech Republic	0.617	25	0.616	0.592	0.642
Malaysia	0.608	26	0.472	0.609	0.606
Hong Kong	0.589	27	0.934	0.604	0.574
Slovenia	0.579	28	0.653	0.579	0.579
China	0.575	29	0.320	0.509	0.640
Saudi Arabia	0.550	30	0.508	0.550	0.550
Spain	0.547	31	0.647	0.554	0.540
Luxembourg	0.544	32	0.928	0.586	0.503
Italy	0.530	33	0.602	0.546	0.513
Chile	0.502	34	0.428	0.494	0.511
India	0.501	35	0.157	0.530	0.472
Estonia	0.497	36	0.730	0.504	0.491
Lithuania	0.485	37	0.625	0.474	0.495
Hungary	0.473	38	0.599	0.430	0.516
Slovak Republic	0.472	39	0.589	0.454	0.491
South Africa	0.472	40	0.249	0.465	0.479
Brazil	0.471	41	0.381	0.492	0.450
Poland	0.468	42	0.514	0.479	0.458
Portugal	0.464	43	0.625	0.475	0.453
Oman	0.454	44	0.379	0.381	0.526
Russian Federation	0.440	45	0.441	0.353	0.527
Tunisia	0.437	46	0.324	0.467	0.407
Indonesia	0.419	47	0.191	0.441	0.399
Thailand	0.402	48	0.313	0.428	0.375
Jordan	0.391	49	0.332	0.400	0.381
Greece	0.379	50	0.517	0.372	0.385
Bahrain	0.375	51	0.700	0.364	0.386
Argentina	0.373	52	0.385	0.380	0.367
Lebanon	0.372	53	0.251	0.441	0.304
Ukraine	0.372	54	0.348	0.328	0.416
Colombia	0.369	55	0.343	0.390	0.348
Turkey	0.368	56	0.389	0.376	0.361
Croatia	0.363	57	0.587	0.337	0.388
Egypt	0.359	58	0.236	0.296	0.421
Kuwait	0.359	59	0.373	0.304	0.413
Romania	0.357	60	0.445	0.337	0.377
Mexico	0.336	61	0.309	0.338	0.334
Kazakhstan	0.327	62	0.338	0.321	0.334
Bulgaria	0.314	63	0.492	0.285	0.342
Philippines	0.305	64	0.224	0.357	0.254
Peru	0.283	65	0.274	0.321	0.245
Iran	0.266	66	0.261	0.269	0.262
Venezuela	0.258	67	0.294	0.209	0.307
Morocco	0.250	68	0.277	0.281	0.219
Mauritania	0.242	69	0.151	0.151	0.332
Algeria	0.233	70	0.182	0.173	0.292
Syria	0.183	71	0.199	0.182	0.185
Libya	0.169	72	0.180	0.115	0.222
Yemen	0.127	73	0.039	0.155	0.099

Table 9: Tacit Macro-Knowledge Competitiveness Indicator (TKCI) and the theme detailed baskets scores, year 2009-2010.

Countries	Tacit Macro-Knowledge Score	Tacit Macro-Knowledge Rank	ICT and E-Services Score	Intellectual Capital Score	Innovation Score
Switzerland	0.955	1	0.953	0.948	0.963
United States	0.944	2	0.838	0.939	0.949
Sweden	0.905	3	0.980	0.910	0.900
Germany	0.891	4	0.928	0.914	0.868
Japan	0.890	5	0.810	0.908	0.872
Denmark	0.869	6	0.951	0.876	0.863
Korea Rep.	0.850	7	0.843	0.809	0.891
Finland	0.837	8	0.826	0.831	0.843
Singapore	0.823	9	0.869	0.816	0.829
United Kingdom	0.821	10	0.917	0.816	0.826
Taiwan China	0.818	11	0.872	0.796	0.840
Austria	0.810	12	0.838	0.827	0.792
Netherlands	0.805	13	0.939	0.837	0.772
Canada	0.764	14	0.845	0.775	0.752
Belgium	0.760	15	0.773	0.776	0.744
France	0.751	16	0.817	0.748	0.753
Norway	0.737	17	0.867	0.739	0.736
Ireland	0.726	18	0.759	0.720	0.733
Israel	0.726	19	0.794	0.683	0.769
Hong Kong	0.709	20	0.969	0.716	0.702
Iceland	0.691	21	0.941	0.671	0.712
Malaysia	0.665	22	0.504	0.689	0.641
Luxembourg	0.662	23	0.935	0.610	0.715
Australia	0.659	24	0.808	0.662	0.655
New Zealand	0.619	25	0.759	0.608	0.631
Czech Republic	0.594	26	0.611	0.608	0.580
Qatar	0.576	27	0.657	0.605	0.547
Italy	0.574	28	0.715	0.594	0.554
United Arab Emirates	0.569	29	0.751	0.628	0.509
Slovenia	0.548	30	0.734	0.556	0.540
India	0.547	31	0.154	0.592	0.502
China	0.536	32	0.365	0.551	0.521
Spain	0.536	33	0.720	0.557	0.514
Estonia	0.523	34	0.789	0.516	0.530
Brazil	0.493	35	0.426	0.540	0.446
South Africa	0.490	36	0.273	0.520	0.460
Indonesia	0.482	37	0.187	0.528	0.436
Tunisia	0.481	38	0.303	0.486	0.477
Kuwait	0.477	39	0.467	0.494	0.460
Portugal	0.476	40	0.677	0.470	0.481
Chile	0.472	41	0.486	0.532	0.411
Saudi Arabia	0.466	42	0.531	0.493	0.438
Slovak Republic	0.463	43	0.615	0.475	0.452
Jordan	0.456	44	0.367	0.450	0.462
Thailand	0.454	45	0.353	0.488	0.420
Lithuania	0.451	46	0.643	0.471	0.430
Hungary	0.448	47	0.610	0.417	0.479
Oman	0.428	48	0.387	0.415	0.441
Turkey	0.425	49	0.440	0.446	0.404
Poland	0.420	50	0.562	0.455	0.385
Philippines	0.382	51	0.253	0.409	0.354
Mexico	0.367	52	0.354	0.400	0.334
Bahrain	0.357	53	0.698	0.360	0.353
Lebanon	0.349	54	0.304	0.395	0.303
Croatia	0.344	55	0.645	0.332	0.355
Ukraine	0.339	56	0.412	0.329	0.350
Greece	0.336	57	0.610	0.356	0.317
Colombia	0.336	58	0.344	0.388	0.283
Romania	0.334	59	0.500	0.335	0.332
Russian Federation	0.327	60	0.527	0.314	0.342
Argentina	0.318	61	0.267	0.349	0.288
Kazakhstan	0.309	62	0.326	0.312	0.307
Morocco	0.309	63	0.286	0.315	0.302
Egypt	0.297	64	0.240	0.319	0.275
Bulgaria	0.282	65	0.528	0.283	0.280
Iran	0.274	66	0.345	0.309	0.239
Peru	0.267	67	0.291	0.321	0.211
Syria	0.238	68	0.265	0.235	0.242
Venezuela	0.171	69	0.343	0.142	0.199
Mauritania	0.149	70	0.134	0.121	0.178
Libya	0.141	71	0.207	0.157	0.124
Algeria	0.131	72	0.463	0.128	0.133
Yemen	0.107	73	0.064	0.135	0.079

References

Abouelnaga, A. E., Metwally, A., Aly, N., Nagy, M., and Agamy, S. 2010. Assessment of nuclear energy sustainability index using fuzzy logic. *Nuclear Engineering and Design*, 240(7):1928 – 1933. 24

Adeodato, P., Arnaud, A. L., Vasconcelos, G., Cunha, R., and Monteiro, D. 2009. Mlp ensembles improve long term prediction accuracy over single networks. *International Journal of Forecasting*, 27(1):661–671. 27, 142, 145

Amin, M. F., Islam, M. M., and Murase, K. 2009. Ensemble of single-layered complex-valued neural networks for classification tasks. *Neurocomputing*, 72(1012):2227 – 2234. 23

Ammar, S., Duncombe, W., Jump, B., and Wright, R. 2004. Constructing a fuzzy-knowledge-based-system: an application for assessing the financial condition of public schools. *Expert Systems with Applications*, 27(3):349–364. 23

Andriantiatsaholiniaina, L. A., Kouikoglou, V. S., and Phillis, Y. A. 2004. Evaluating strategies for sustainable development: fuzzy logic reasoning and sensitivity analysis. *Ecological Economics*, 48(2):149 – 172. 24

Archibugi, D., Denni, M., and Filippetti, A. 2009. The technological capabilities of nations: The state of the art of synthetic indicators. *Technological Forecasting and Social Change*, 76(7):917–931. 3, 19, 42

Arndt, C. and Oman, C. 2006. Uses and abuses of governance indicators. *OECD Development Centre*. 14

Atiya, A. F., M., E.-S. S., I., S. S., and S., E.-S. M. 1999. A comparison between neural-network forecasting techniques - case study: River flow forecasting. *IEEE Transactions on Neural Networks*, 10(2). 145

Babbie, E. R. 1995. *The practice of social research*. Wadsworth Publishing Company. 18

Baker, Bruce D.; Richards, C. E. 1999. A comparison of conventional linear regression methods and neural networks for forecasting educational spending. *Economics of Education Review*, 18(4):405–515. 143

Baltagi, H. B. 2005. *Econometric Analysis of Panel Data*, volume 3. John Wiley & Sons, New York, USA. 87, 88

Baltagi, H. B. and Li, Q. 1992. Prediction in the one - error component model with serial correlation. *Journal of Forecasting*, 11(1):561–567. 88

Bektas Ekici, B. and Aksoy, U. T. 2011. Prediction of building energy needs in early stage of design by using anfis. *Expert Systems with Applications*, 38(5):5352–5358. 27

Belhadj, B. 2012. New weighting scheme for the dimensions in multidimensional poverty indices. *Economics Letters*, 116(3):304 – 307. 18

Bezdek, J. C., Ehrlich, R., and Full, W. 1984. Fcm: The fuzzy c-means clustering algorithm. *Computers and Geosciences*, 10(2-3):191–203. 79

Bhanojirao, V. 1991. Human development report 1990: Review and assessment. *World Development*, 19(10):1451–1460. 18

Bontis, N., Dragonetti, N. C., Jacobsen, K., and Roos, G. 1999. The knowledge toolbox:: A review of the tools available to measure and manage intangible resources. *European Management Journal*, 17(4):391–402. 2

Botha, A. P. 2007. *Knowledge: Living and Working with it*, volume 1. Juta Legal and Acedemic Publishers. 4

Brooking, A. 1996. *Intellectual capital*. Thomson Learning. 34

Centre, E. C. J. R. 2012. Composite indicators and ranking systems. http://composite-indicators.jrc.ec.europa.eu [Online; accessed: 17/04/2012]. 15

Cerulli, G. and Filippetti, A. 2012. The complementary nature of technological capabilities: Measurement and robustness issues. *Technological Forecasting and Social Change*, 79(5):875 – 887. 18

Chandra, K. e. 2010. Constructivist theories of ethnic politics. http://politics.as.nyu.edu/docs/IO/2587/Chapter1-Introduction-09-28-10.pdf [Online; accessed:15/08/2012]. 36

Charles, V. 2013. Competitiveness evaluation methodologies, contemporary issues and challenges. *Journal of CENTRUM Cathedra*, 6(1). 38

Chen, D. and Dahlman, C. 2005. The knowledge economy, the kam methodology and world bank operations. *World Bank Institute Working Paper*, (37256). 33, 57

Chen, D. and Kee, H. 2005. A model on knowledge and endogenous growth. *World Bank Policy Research Working Paper*, (3539). 33, 57

Chen, M.-Y., Huang, M.-J., and Cheng, Y.-C. 2009. Measuring knowledge management performance using a competitive perspective: An empirical study. *Expert Systems with Applications: An International Journal*, 36(4):120–120. 137

Cheng, P., Quek, C., and Mah, M. 2007. Predicting the impact of anticipatory action on us stock marketan event study using anfis (a neural fuzzy model). *Computational Intelligence*, 23(2):117–141. 28

Cherchye, L. and Kuosmanen, T. 2004. Benchmarking sustainable development: a synthetic meta-index approach. Technical report, Research Paper, UNU-WIDER, United Nations University (UNU). 6, 15, 19, 46, 116

Cherchye, L., Lovell, C. K., Moesen, W., and Puyenbroeck, T. V. 2007a. One market, one number? a composite indicator assessment of eu internal market dynamics. *European Economic Review*, 51(3):749 – 779. 18, 55

Cherchye, L., Moesen, W., and Puyenbroeck, T. 2004. Legitimately diverse, yet comparable: On synthesizing social inclusion performance in the eu. *JCMS: Journal of Common Market Studies*, 42(5):919–955. 18

Cherchye, L., Moesen, W., Rogge, N., Van Puyenbroeck, T., Saisana, M., Saltelli, A., Liska, R., and Tarantola, S. 2007b. Creating composite indicators with dea and robustness analysis: the case of the technology achievement index. *Journal of the Operational Research Society*, 59(2):239–251. 18

Choo, C. W. 1998. *The knowing organization: How organizations use information to construct meaning, create knowledge, and make decisions*, volume 1. Oxford University Press. 31

Conceio, P., Gibson, D. V., Heitor, M. V., and Sirilli, G. 2001. Beyond the digital economy: A perspective on innovation for the learning society. *Technological Forecasting and Social Change*, 67(23):115 – 142. 33

Conceio, P., Heitor, M. V., and Oliveira, P. M. 1998. Expectations for the university in the knowledge-based economy. *Technological Forecasting and Social Change*, 58(3):203 – 214. 61

Cortright, J. 2001. New growth theory, technology and learning:a practitioners guide. 32

Coupland, S. and John, R. 2008. New geometric inference techniques for type-2 fuzzy sets. *International Journal of Approximate Reasoning*, 49(1):198 – 211. 163

Crone, S. F., Hibon, M., and Nikolopoulos, K. 2011. Advances in forecasting with neural networks? empirical evidence. *International Journal of Forecasting*, 27(1):635–660. 27, 142

Cumbrera, G., Almenara-Barrios, J., Lopez-Lara, Peralta-Saez, and Carulla, G. 2008. Development and spatial representation of synthetic indexes... *Epidemiologia e Psichiatria Sociale*, 17(3):55–64. 15

Dakka, W. and Ipeirotis, P. G. 2008. Automatic extraction of useful facet hierarchies from text databases. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, pages 466–475. IEEE. 47

Dempester, A. P., Laird, N. M., and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society*, 39(1):1 – 38. 85

Dostál, P. 1998. Some knowledge of prediction methods. *Nostradamus Prediction Workshop*, (6). 137

Dostál, P. 2009. Stock market decision making machine. *29th International Symposium on Forecasting*, (1):171,5. 24

Drucker, P. F. 1957. *Landmarks of Tomorrow*. Harper and Row. 32

Drumond, L. and Girardi, R. 2010. Extracting ontology concept hierarchies from text using markov logic. In *Proceedings of ACM Symposium on Applied Computing*, pages 1354–1358. ACM. 47

Dunham, M. H. 2006. *Data mining: Introductory and advanced topics*. Pearson Education India. 48

Dunn, J. C. 1973. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57. 79

Dutta, S. and Bilbao-Osori, B. 2012. The global information technology report - living in a hyperconnected world. http://www.weforum.org/gitr [Online; accessed 02/05/2012]. 51

Dyba, T. 2000. An instrument for measuring the key factors of success in software process improvement. *Empirical Software Engineering*, (1):357 – 390. 17

Ebert, U. and Welsch, H. 2004. Meaningful environmental indices: a social choice approach. *Journal of Environmental Economics and Management*, 47(2):270 – 283. 22

EuropeanCommission 2011. Innovation union scoreboard. *Pro Inno Europe*, 1(1):101. 18, 22

Fausett, L. 1994. *Fundamentals of Neural Networks*, volume 1. Prentice Hall, Inc., New Jersy, USA. 90, 144

Foray, D. and Lundvall, B.-A. 1996. The knowledge-based economy: from the economics of knowledge to the learning economy. *Employment and Growth in the Knowledge-based Economy*, pages 11–32. 14

Freudenberg, M. 2003. Composite indicators of country performance: a critical assessment. Technical report, OECD Publishing. 42, 43, 44, 74, 103, 137

Ghaseminezhad, M. H. and Karami, A. 2011. A novel self-organizing map (som) neural network for discrete groups of data clustering. *Applied Soft Computing*, 11(6):3771–3778. 92, 144

Girma, S. 2008. Econometrics/ lecture handouts. 87

Goldman, A. 1999. *Knowledge in a Social World*. Oxford:Clarendon Press. 32

Gordon, T. J. 2005. State of the future index (sofi) method. *AC/UNU Millennium Project: Futures Research Methods-V2*. 137

Gray, R. 1984. Vector quantization. *ASSP Magazine, IEEE*, 1(2):4–29. 82

Grupp, H. and Schubert, T. 2010. Review and new evidence on composite innovation indicators for evaluating national performance. *Research Policy*, 39(1):67 – 78. 17, 18, 78

Hatefi, S. and Torabi, S. 2010. A common weight mcda dea approach to construct composite indicators. *Ecological Economics*, 70(1):114 – 120. 18

Hathaway, R. J. and Bezdek, J. C. 2001. Fuzzy c-means clustring of incomplete data. *IEEE Transactions on Systems, Man, and Cybernetics*, 31(5):735 – 744. 24, 83

Herrero, A., Corchado, E., and Jimenez, A. 2011. Unsupervised neural models for country and political risk analysis. *Expert Systems with Applications*, 38(11):13641 – 13661. 25, 27

Hetherington, S. 2012. Knowledge. http://www.iep.utm.edu/knowledg/ [Online; accessed 12/01/2013]. 29

Hopgood, A. A. 2011. *Intelligent systems for engineers and scientists*. CRC PressI Llc. 94, 95

Hopkins, M. 1991. Human development revisited: A new undp report. *World Development*, 19(10):1469–1473. 18

Hsiao, C. 2003. *Analysis of Panel Data*, volume 2. Cambridge University Press, New York, USA. 88, 92, 141, 142, 144

Huberman, B. A., Pirolli, P. L., Pitkow, J. E., and Lukose, R. M. 1998. Strong regularities in world wide web surfing. *Science*, 280(5360):95–97. 47

Hwang, C. and Rhee, F. C.-H. 2007. Uncertain fuzzy clustering: interval type-2 fuzzy approach to c-means. *Fuzzy Systems, IEEE Transactions on*, 15(1):107–120. 163

Ilonen, J., Kamarainen, J.-K., Puumalainen, K., Sundqvist, S., and Klviinen, H. 2006. Toward automatic forecasts for diffusion of innovations. *Technological Forecasting and Social Change*, 73(02):182–198. 26, 27, 92, 137, 142, 144

IMD 2011. World competitiveness yearbook. Technical Report 2011, International Institute for Management Development. http://www.imd.org/research/publications/wcy/index.cfm [Online; accessed 10/06/2011]. 51, 69

INSEAD 2011. Global innovation index. 51

ITU 2012. Measuring the information society, the ict development index. Technical Report 2012, International Telecommunications Union. http://www.itu.int/pub/D-IND-ICTOI-2012/en [Online; accessed 26/09/2012]. 17, 51, 59

Jammazi, R. and Aloui, C. 2012. Crude oil price forecasting: Experimental evidence from wavelet decomposition and neural network modeling. *Energy Economics*, 34(3):828 – 841. 104

Jang, J.-S. 1993. Anfis: adaptive-network-based fuzzy inference system. *Systems, Man and Cybernetics, IEEE Transactions on*, 23(3):665–685. 93

Johnson, D. M., Porter, A. L., Roessner, D., Newman, N. C., and Jin, X.-Y. 2010. High-tech indicators: assessing the competitiveness of selected european countries. *Technology Analysis & Strategic Management*, 22(3):277–296. 101

Johnson, R. A. and Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis*, volume 6. Pearson. 78, 109

Jolliffe, I. T. 2002. *Principal component analysis*. Springer verlag. 73

Karnik, N. N., Mendel, J. M., and Liang, Q. 1999. Type-2 fuzzy logic systems. *Fuzzy Systems, IEEE Transactions on*, 7(6):643–658. 163

Keles, A., Kolcak, M., and Keles, A. 2008. The adaptive neuro-fuzzy model for forecasting the domestic debt. *Knowledge-Based Systems*, 21(8):951–957. 28, 95

Keller, A. A. 2008. Fuzzy control of macroeconomic models. *World Academy of Science, Engineering and Technology*, 41(41):1307–6884–1307–6884. 24

Kennedy, P. 2003. *A guide to econometrics*. MIT press. 93

Kershaw, P. and Rossini, P. 1999. Using neural networks to estimate constant quality house price indices. In *Pacific-Rim Real Estate Society Conference*, page 224, Malaysia. Pacific-Rim Real Estate Society Conference. 24, 145

Khun, T. S. 1996. *The Structure of Scientific Revolutions*, volume 3. University of Chicago Press. 35, 46

Kietz, J.-U., Maedche, A., and Volz, R. 2000. A method for semi-automatic ontology acquisition from a corporate intranet. In *Workshop Ontologies and text*. 47

Klein, P. D. 1998, 2005. Epistemology. *Routledge Encyclopedia of Philosophy*. Retrieved December 28, 2012. 29

Kohonen, T. 1990. The self-organizing map. In *IEEE Xplore*, volume 78, pages 1464 –1480. 90

Korotayev, V. A. and Tsirel, V. S. 2010. A spectral analysis of world gdp dynamics: Cycles in global economic development, and the 2008 / 2009 economic crisis. *The Structure and Dynamics*, 4(1). 142

Kumiega, A. and Vliet, B. V. 2008. Perform in-sample/out-of-sample tests. In *Quality Money Management*, pages 157 – 163. Academic Press, Burlington. 104

Lambsdorff, J. G. 2003. Background paper to the 2003 corruption perceptions index. *Manuscript, Transparency International, Passau, Ger.* 62

Linde, Y., Buzo, A., and Gray, R. 1980. An algorithm for vector quantizer design. *Communications, IEEE Transactions on*, 28(1):84–95. 81, 82, 165

Little, R. J. A. and Rubin, D. B. 2002. *Statistical Analysis with Missing Data - Second Edition*. Wiley-Interscience, New Jersey, USA. 85

Liu, K. 2007. Evaluating environmental sustainability: An integration of multiple-criteria decision-making and fuzzy logic. *Environmental Management*, 39(5):721–736. 25

Lucas, R. E. 1988. On the mechanics of economic development. *Journal of monetary economics*, 22(1):3–42. 60

Lustick, I. S. 2000. Agent-based modelling of collective identity: testing constructivist theory. *Journal of Artificial Societies and Social Simulation*, 3(1). 36

M. Saisana, A. S. and Tarantola, S. 2005. Composite indicators - the controversy and the way forward. *Statistics, Knowledge and Policy, OECD World Forum on Key Indicators*, 1(1):2–4. 14, 43, 44

Maddala, K. L. G. 2009. *Introduction to econometrics*. Wiley, UK. 88

Maesschalck, R. D., Jouan-Rimbaud, D., and Massart, D. 2000. The mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50(1):1 – 18. 101

Mahlberg, B. and Obersteiner, M. 2001. Remeasuring the hdi by data envelopement analysis. *Interim Reports on work of the International Institute for Applied Systems Analysis*, (1):36. 19

Makridakis, S., Wheelwright, S. C., and Hyndman, R. J. 1998. *Forecasting methods and applications*. John Wiley & Sons, Inc., New York, USA. 26

Mantel, N. 1963. Chi-square tests with one degree of freedom; extensions of the mantel-haenszel procedure. *Journal of the American Statistical Association*, 58(303):690–700. 76

Markus, G. B. 1979. *Analyzing Panel Data*, volume 1. Sage University Paper, California, USA. 87

Mattelart, A. 2003. *The Information Society.* SAGE Publications Ltd. 32

Meade, N. and Islam, T. 2006. Modelling and forecasting the diffusion of innovation – a 25-year review. *International Journal of Forecasting*, 22(3):519–545. 142

Mendel, J. M. and John, R. B. 2002. Type-2 fuzzy sets made simple. *Fuzzy Systems, IEEE Transactions on*, 10(2):117–127. 163

Mendel, J. M., John, R. I., and Liu, F. 2006. Interval type-2 fuzzy logic systems made simple. *Fuzzy Systems, IEEE Transactions on*, 14(6):808–821. 163

Merlin, P., Sorjamaa, A., Maillet, B., and Lendasse, A. 2010. X-som and l-som: A double cassfication approach for missing value imputation. *Neurocomputing*, 73(1):1103–1108. 90

Michael, F. N. R. G. 2010. Understanding versus knowledge the new role of ict and elearning. http://www.nexusresearchgroup.com/information-systems/article1.htm [Online; accessed 20/11/2009]. 2

Mitchell, G., May, A., and McDonald, A. 1995. A methodological framework for the development of indicators of sustainable development. *The International Journal of Sustainable Development & World Ecology*, 2(2):104–123. 21

Munda, G. and Nardo, M. 2005. Constructing consistent composite indicators: the issue of weights. *EUR 21834 EN.* 22

Munda, G. and Nardo, M. 2009. Noncompensatory/nonlinear composite indicators for ranking countries: a defensible setting. *Applied Economics*, 41(12):1513–1523. 22

Murray, R., Harding, D., Angus, T., Gillespie, R., and Arora, H. 2012. Emigration from the uk. http://www.homeoffice.gov.uk/publications/science-research-statistics/research-statistics/immigration-asylum-research/horr68/horr68-report?view=Binary [Online; accessed 12/01/ 2013]. 32

Nardo, M., Saisana, M., Saltelli, A., and Tarantola, S. 2005. Tools for composite indicators building. *European Comission, Ispra*. 15, 19, 20

Natoli, R. and Zuhair, S. 2010. What is a reasonable measure of progress? *Emerald: International Journal of Sociology and Social Policy*, 30(1):201 – 218. 36

Neal, R. M. and Hinton, G. E. 1998. A view of the em algorithm that justifies incremental, sparse, and other variants. In *Learning in graphical models*, pages 355–368. Springer. 85

Nicoletti, G., Scarpetta, S., and Boylaud, O. 1999. Summary indicators of product market regulation with an extension to employment protection legislation. 110

Niwa, F. and Tomizawa, H. 1996. A trial of general indicator of science and technology: Methodological study of overall estimation of national s&t activity. *Scientometrics*, 37:245–265. 16

Nonaka, I. 1994. A dynamic theory of organizational knowledge creation. *Orgnization Science*, 5:14 – 37. 31

Nuovo, A. G. D. 2011. Missing data analysis with fuzzy c-means: A study of its application in a psychological scenario. *Expert Systems with Applications*, 38(6):6793 – 6797. 24

Nussbaumer, P., Bazilian, M., and Modi, V. 2012. Measuring energy poverty: Focusing on what matters. *Renewable and Sustainable Energy Reviews*, 16(1):231 – 243. 23

OECD 2002. International mobility of the highly skilled. http://depts.washington.edu/gcp/pdf/globalcitizenship.pdf [Online; accessed 12/01/ 2013]. 32

OECD 2008a. Handbook on constructing composite indicators: Methodology and user guide. http://www.oecd.org/dataoecd/37/42/42495745.pdf [Online accessed 17/02/2012]. 18, 22, 25, 43, 44, 72, 77, 78, 107

OECD 2008b. Statistics, knowledge and policy 2007 measuring and fostering the progress of societies. *OECD Papers*. 5, 14

Osberg, L. and Sharpe, A. 2002. International comparisons of trends in economic well-being. In *Assessing Quality of Life and Living Conditions to Guide National Policy*, volume 11 of *Social Indicators Research Series*, pages 349–382. Springer Netherlands. 4, 20

Osberg, L. and Sharpe, A. 2005. How should we measure the economic aspects of well-being? *Review of Income and Wealth*, 51(2):311–336. 4

OxfordEconomics 2011. The new digital economy. http://www.pwc.com/gx/en/technology/publications/assets/the-new-digital-economy.pdf [Online; accessed 9/01/2013]. 34

Panigrahi, R. and Sivramkrishna, S. 2002. An adjusted human development index: robust country rankings with respect to the choice of fixed maximum and minimum indicator values. *Journal of Human Development*, 3(2):301–311. 18

Pao, H.-T. and Chih, Y.-Y. 2006. Comparison of tscs regression and neural network models for panel data forecasting: debt policy. *Neural Computing & Applications*, 15(2):117–123. 25

Phillis, Y. A., Grigoroudis, E., and Kouikoglou, V. S. 2011. Sustainability ranking and improvement of countries. *Ecological Economics*, 70(3):542–553. 24

Polanyi, M. 1962. *Personal Knowledge: Towards a Post-Critical Philosophy*, volume 3. University Of Chicago Press. 30

Polanyi, M. 1967. The tacit dimension. 31

Polanyi, M. 2009. *The Tacit Dimension*. University of Chicago press. 31

Porter, A. L., Newman, N. C., Roessner, J. D., Johnson, D. M., and Jin, X.-Y. 2009. International high tech competitiveness: does china rank number 1. *Technology Analysis & Strategic Management*, 21(2):173–193. 101

Porter, M. E. 1998. *The Competitive Advantage of Nations*, volume 1. Palgrave Macmillan, New York, USA. 37

Porter, M. E. and Stern, S. 1999. National innovation capacity index. 19

Quantumiii 2011. Competitive intelligence glossary. [Online; accessed 02/12/2009]. 2, 32

Ramamurthi, B. and Gersho, A. 1986. Classified vector quantization of images. *Communications, IEEE Transactions on*, 34(11):1105–1115. 82

Rhee, F. C.-H. and Hwang, C. 2001. A type-2 fuzzy c-means clustering algorithm. In *IFSA World Congress and 20th NAFIPS International Conference, 2001. Joint 9th*, volume 4, pages 1926–1929. IEEE. 163

Riley, T. B. 2003. Knowledge management and technology. *International tracking survey report*, 3. 2

Robertson, S. L. 2009. Market multilateralism, the world bank group and the asymmetries of globalising higher education: Toward a critical political economy analysis. *International organizations and higher education policy: Thinking globally, acting locally*, pages 113–131. 33

Roessner, J. D., Porter, A. L., Newman, N., and Cauffiel, D. 1996. Anticipating the future high-tech competitiveness of nations: indicators for twenty-eight countries. *Technological Forecasting and Social Change*, 51(2):133–149. 101

Romer, P. 1991a. Endogenous technological change. Technical report, National Bureau of Economic Research. 60

Romer, P. M. 1989. Human capital and growth: theory and evidence. Technical report, National Bureau of Economic Research. 32, 56

Romer, P. M. 1991b. Endogenous technological change. Technical report, National Bureau of Economic Research. 33, 56

Rosen, R. 2005. *Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*, volume 1. Columbia University Press, New York, USA. 39

Rouse, M. 2007. Intellectual capital. Retrieved January 10, 2013. 35

Saaty, T. L. 2008. Decision making with the analytic hierarchy process. *International Journal of Services Sciences*, 1(1):83. 22

Saisana, M. and Munda, G. 2008. *Knowledge Economy: measures and drivers*. JRC Scientific and Technical Reports. 6, 25, 46

Saltelli, A., Chan, K., Scott, E. M., et al. 2000. *Sensitivity analysis*, volume 134. Wiley New York. 93

Saltelli, A., Saisana, M., Nardo, M., and Annoni, P., editors 2012. *Composite Indicators and Ranking Systems: Construction and Use*. Wiley-Blackwell. 5, 14, 15, 21

Sandhusen, R. L. 2008. *Marketing*, volume 4. Barron's Educational Series. 46

Sarle, W. S. 1994. Neural networks and statistical models. 74, 138

Sarlin, P. 2010. Visual monitoring of financial stability with a self-organizing neural network. In *10th IEEE International Conference on Intelligent Systems Design and Applications*, pages 1 –8. 23, 27

Sarlin, P. 2011. Sovereign debt monitor: A visual self-organizing maps approach. In *Computational Intelligence for Financial Engineering and Economics (CIFEr), 2011 IEEE Symposium on*, pages 1 –8. 27, 28, 92

SAS.com 2011. Principal component analysis. http://support.sas.com/publishing/pubcat/chaps/55129.pdf [Online accessed 17/11/2011]. 147

Schafer, J. 1997. *Analysis of Incomplete Multivariate Data*, volume 1. Chapman and Hall/CRC. 85

Self, A., Thomas, J., Randall, C., et al. 2012. Measuring national well-being: Life in the uk, 2012. 4

Sharpe, A. and Andrews, B. 2012. An assessment of weighting methodologies for composite indicators: The case of the index of economic well-being (iewb). http://www.csls.ca/reports/csls2012-10.pdf [Retrieved January 5, 2013]. 20

Shnaider, E. and Haruvy, N. 2008. Background factors facilitating economic growth using linear regression and soft regression. *Fuzzy Economic Review*, xiii(1):41–55–41–55. 23, 24

Smith, L. I. 2002. A tutorial on principal components analysis. 27

Spender, J. C. 1996. Making knowledge the basis of a dynamic theory of the firm. *Strategic Management Journal*, 17:45 – 62. 31

Spender, J.-C. 1998. Pluralist epistemology and the knowledge-based theory of the firm. *Organization*, 5(2):233–256. 30, 31

Stack, J. F. 1986. *"Regional Economic Development and the Territorial Basis of Ethnicity" in The Primordial Challenge: Ethnicity in the Contemporary World*, volume 1. Greenwood Press, CT, USA. 35

Stair, R. and Reynolds, G. 2007. *Fundamentals of Information Systems*, volume 1. Course Technology, Boston, USA, 6 edition. 2

Statistics.gov.uk 2010. Indicators. www.neighbourhood.statistics.gov.uk [Online; accessed 02/12/2010]. 14

Stenmark, D. 2001. The relationship between information and knowledge. In *Proceedings of IRIS*, volume 24, pages 11–14. 30

Stevens, C. 1996. The knowledge-driven economy. *OECD observer*, 200(1):6–10. 60

Stewart, G. M. 2007. Kaupapa mori science. Retrieved February 1, 2012. 30

Surry, D. and Farquhar, J. 1997. Diffusion theory and instructional technology. *Journal of Instructional Science and Technology*, 2(1):24–36. 35

Tarantola, S., Nardo, M., Saisana, M., and Gatelli, D. 2006. A new estimator for sensitivity analysis of model output: An application to the e-business readiness composite indicator. *Reliability Engineering and System Safety*, 91(10–11):1135 – 1141. 5, 44

Teece, D. 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research policy*, 15(6):285–305. 34

Thang, K. F., Aggarwal, R. K., McGrail, A. J., and Esp, D. 2003. Analysis of power transformer dissolved gas data using the self-organizing map. *IEEE Transactions on Power Delivery*, 18(4):1241–1248. 92, 144

Trebilcock, M. and Prado, M. 2011. *What Makes Poor Countries Poor: Institutional Determinants of Development*, volume 1. Edward ElgarPublishing Limited, Cheltenham,UK and Massachusetts, USA. 5, 44

UN, E. W. G. o. S. E. 2010. Measuring the knowledge-based economy: A statistical framework for measuring knowledge in the australian economy and society. 14

Urra, F.-J. 2007. Assessing corruption an analytical review of corruption measurement and its problems: Perception, error and utility. *Edmund A. Walsh School of Foreign Service, Georgetown University*. Online; accessed: 19/11/2012. 14

WB 2010. Knwoledge for development: Kam tool. Technical Report 2009-2010, World Bank. www.worldbank.org/kam [Online; accessed 26/04/2011]. 51

WEF 2011. Global competitiveness report. Technical Report 2011, World Economic Forum. http://www.weforum.org/reports [Online; accessed: 20/07/2012]. 51

West, D. and Dellana, S. 2011. An empiricial analysis of neural network memory structures for basin water quality forecasting. *International Journal of Forecasting*, 27(1):777–803. 26

Willmott, C., Ackleson, S., Davis, R., Feddema, J., Klink, K., Legates, D., O'Donnell, J., and Rowe, C. 1985. Statistics for the evaluation and comparison of models. *Journal of Geophysical Research*, 90(c5):8995–9005. 104

Wilson, I. D., Paris, S. D., Ware, J. A., and Jenkins, D. H. 2002. Residential property price time series forecasting with neural networks. *Knowledge-Based Systems*, 15(5-6):335–341. 23, 27, 89, 143

Wu, X. and Guan, L. 1994. Acceleration of the lbg algorithm. *Communications, IEEE Transactions on*, 42(234):1518–1523. 82

Yadav, V. and Srinivasan, D. 2011. A som-based hybrid linear-neural model for short-term load forecasting. *Neurocomputing*, 74(17):2874 – 2885. 27

Yaffee, R. 2003. *A Primer for Panel Data Analysis*, volume 1. New York University Press, New York, USA. 87, 88

Yale 2012. Environmental performance index. Retrieved January 7, 2013. 17

Zhao, X. and Wei, G. 2013. Some intuitionistic fuzzy einstein hybrid aggregation operators and their application to multiple attribute decision making. *Knowledge-Based Systems*, 37(0):472 – 479. 25, 107

Zheng, Y., Julstrom, B., and Cheng, W. 1997. Design of vector quantization codebooks using a genetic algorithm. In *Evolutionary Computation, 1997., IEEE International Conference on*, pages 525–529. IEEE. 81

Zhou, P., Ang, B., and Poh, K. 2006. Comparing aggregating methods for constructing the composite environmental index: An objective measure. *Ecological Economics*, 59(3):305–311. 22

Zhu, B., Xu, Z., and Xia, M. 2012. Hesitant fuzzy geometric bonferroni means. *Information Sciences*, 205(0):72 – 85. 25, 107

Zimmermann, H. G. 2011. Advances in time series modelling and forecasting with neural networks. *ISF 2011, Workshop Manual*. 92, 144

Zimmermann, H. J. and Zysno, P. 1983. Decisions and evaluations by hierarchical aggregation of information. *Fuzzy Sets and Systems*, 10(6):243–260. 79, 110