 Unified knowledge based economy hybrid forecasting

This item was submitted to Loughborough University's Institutional Repository by the/an author.

**Citation:** AL-SHAMII, A. ...et al., 2015. Unified knowledge based economy hybrid forecasting. Technological Forecasting and Social Change, 91, pp.107-123.

**Additional Information:**

- This paper was accepted for publication in the journal Technological Forecasting and Social Change and the definitive published version is available at: http://dx.doi.org/10.1016/j.techfore.2014.01.014

**Metadata Record:** [https://dspace.lboro.ac.uk/2134/22865](https://dspace.lboro.ac.uk/2134/22865)

**Version:** Accepted version

**Publisher:** © Elsevier

**Rights:** This work is made available according to the conditions of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence. Full details of this licence are available at: https://creativecommons.org/licenses/by-nc-nd/4.0/

Please cite the published version.
Unified Knowledge Based Economy Forecasting

Ahmad Al Shami*, Ahmad Lotfi*, Simeon Coleman†, Petr Dostál‡

* School of Science and Technology, Nottingham Trent University
  Nottingham, United Kingdom
† Nottingham Business School, Nottingham Trent University
  Nottingham, United Kingdom
‡ Faculty of Business and Management, Institute of Informatics
  Brno University of Technology, Brno, Czech Republic

Abstract

In this paper, a new framework for forecasting Knowledge Based Economy (KBE) competitiveness is proposed. Existing KBE indicators from internationally recognised organisations including the World Bank and the International Telecommunication Union among others are used to forecast and unify the KBE performance indices. Three different forecasting methods including Panel Data: time-series cross sectional (TSCS), Linear Multiple Regression (LMREG), and Artificial Neural Network (ANN) are investigated. The ANN forecasting model outperformed the TSCS and LMREG. The proposed KBE forecasting model utilizes a 2-stage ANN model which are 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. Feed-forward neural network is used to learn and predict the scores of nations using past observed data. Then, SOM is used to aggregate the forecasted scores and to place nations in homogeneous clusters. The proposed framework can be applied in the context of forecasting the competitiveness of KBE for any nation even over small time periods or missing data.

Keywords: Knowledge Based Economy; Strategic Forecasting; Computational Intelligence; Artificial Neural Network (ANN); Self-Organizing Map (SOM); Time-Series Cross Sectional (TSCS); Multiple Regression.
1. Introduction

Utilising knowledge and innovation as the key driver for prosperity and growth on a macro level leads to the creation of Knowledge Based Economy (KBE) or Knowledge Economy (KE). KBE relies on the production, distribution, and use of knowledge and innovation as the main driver of growth, wealth creation, and employment across all industries. This means that a nation must be in a constant process of change and adaptation to the new economic realities. It is evident that the key to success is the diffusion of knowledge on a macro scale. For governments they have to invest in, and develop, knowledge workers to create or maintain a knowledge society [23].

Many synthetic indicators have been created by highly regarded organisations including the World Bank (WB), the World Economic Forum (WEF), the United Nation (UN), the International Institute for Management Development (IMD), and the International Telecommunication Union (ITU). These indicators have been used by organisations including government agencies, aid agencies and research institutions to assess the competitiveness of a nation or nations in the context of KBE. However, these indicators yield different scores and ranking depending on the nature and type of assessments. Moreover, all of these indicators report past performance, and it does not predict where a certain KBE is heading giving all known elements.

Measuring and predicting the directions of KBE and innovation on a macro scale and competitive level has gained momentum. A similar concept is the primary concern of the “State of the Future Index” [9], 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 [11]. A study by Chen et al. [5] 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 [6].

Many contemporary studies presented computational intelligence techniques to deal with the prediction of complicated and uncertain cases, for example Ilonen et al. [13] integrated
Bass model, Artificial Neural Network (ANN), and Kohonen Self-Organizing Map (SOM) to forecast the diffusion of innovations in nations. Recent study by Lee et al. [20] 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. Wilson et al. [27] presented an ANN model to forecast the future trends within any economy residential property. Studies by Dostál [7] and Kershaw and Rossini [17], incorporated hybrid methods including 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.

As an alternative to computational intelligence techniques, Panel Data or Time-Series Cross-Sectional (TSCS) analysis are usually used to estimate parameters of an economical system. A unique study by Pao and Chih [22] concluded that ANN models can be used to solve panel data regression, and that would allow us to construct and test sophisticated models than purely cross-sectional or time-series data to solve debt policy forecasting problems.

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

The remainder of this paper is organised as follows. In Section 2, the KBE progress forecasting framework is presented and in Section 3, the data sources and the indicators selection for the study are explained. Section 4 includes the proposed forecasting methodologies for the multiple regression, panel data analysis and the ANN based KBE forecasting models, followed by performance measures, tests, analysis and the results of our findings in Section 5. The conclusions and future research are summed up in Section 6.
2. Knowledge Based Economy Forecasting Framework

When the progress and competitiveness of KBE in a specific economy is forecasted, 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 five major KBE related indices is shown in Figure 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 related macro-knowledge synthetic aggregated indices for the forecasting models under consideration. The
next step down is to test whether the selected indicators have a relationship, and how strong? If the indicators selected have no relationship or a very weak one, it is recommended to search for new macro-knowledge related indicators. If there is a significant relationship between the indicators, then the standardization process like normalisation will begin. If, however, the indicators are standardised or normalised, then the next step would be to carry on with the forecasting process. The forecasting step is the core of this framework and in this step three major forecasting techniques will be utilized; 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 winner performance results will be clustered using Kohonen’s Self-Organizing Map, then visualised and evaluated. The framework exhibited in Figure 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.

3. Data Source and Descriptions

Many individual and analogous composite indicators are developed to measure all aspects of progress and developments of a nation on a micro and macro level, and it is incomprehensible to give in-depth analysis to all of them. The focus will be on selection of five indicators, according to their relevance, availability and wide usage. Collected data set has been sampled from the freely and readily available annual reports of the following indices:

- ICT Development Index (IDI), developed by the International Telecommunication Union [16]
- Global Competitiveness Index (GCI) from the World Economic Forum [26]
- INSEAD Global Innovation Index (GII) [15]
Table 1: Summary of the main Knowledge Economy indicators.

<table>
<thead>
<tr>
<th>Indicator</th>
<th>ITU(IDI)</th>
<th>WB(KEI)</th>
<th>WEF(GCR)</th>
<th>IMD(WCY)</th>
<th>INS(GII)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Range</td>
<td>0 - 10</td>
<td>0 - 10</td>
<td>1 - 7</td>
<td>0 - 100</td>
<td>0 - 100</td>
</tr>
<tr>
<td>Countries</td>
<td>159</td>
<td>146</td>
<td>134</td>
<td>57</td>
<td>125</td>
</tr>
</tbody>
</table>

- World Competitiveness Yearbook (WCY) from the International Institute for Management Development [14] and

- Knowledge Economy Index (KEI) developed by the World Bank [25].

The period that these KBE indices are generated vary, and even the range and economies (countries) included in these indices could be different. In Table 1 a summary of the availability of the main KBE indicators is presented. It should be noted that no scores or ranking were issued by the World Bank after 2009, and the INS (GII) report edition was first released in 2009. To make the research comparable across all indices, collected data set was obtained by matching the overall reported scores for 51 economies over five years period from 2007 to 2011 from ITU, WEF and IMD indices. In addition to these three, the recent available scores from WB for three years periods from 2007 to 2009 and from INS for the same set of economies from 2009 to 2011 were matched and used in this study.

3.1. Correlation Analysis

Correlation analysis are conducted to identify the relationship between the selected KBE indicators. The analysis was conducted using the full available dataset. Correlation coefficient matrix is summarised in Table 2. The results show highest correlation between the WEF and INS indicators with 0.96 correlation coefficient and between ITU and WB indicators with 0.95 correlation coefficient. The results also showed a high to moderate correlation that ranges from 0.74 to 0.76 between WEF and ITU, WEF and WB, INS and ITU and also INS and WB indicators. The lowest correlation was moderate which resulted between the ITU and IMD with 0.67 and WB and IMD with 0.67 correlation coefficient. Based on these results it can be confirm that the selected indicators are related and comparable.
Table 2: Correlation coefficient matrix for all selected indices.

<table>
<thead>
<tr>
<th>51-Economies</th>
<th>ITU-IDI</th>
<th>WB-KEI</th>
<th>WEF-GCI</th>
<th>IMD-WCY</th>
<th>INS-GII</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITU-IDI</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WB-KEI</td>
<td>0.950</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEF-GCI</td>
<td>0.743</td>
<td>0.757</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMD-WCY</td>
<td>0.676</td>
<td>0.673</td>
<td>0.884</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>INS-GII</td>
<td>0.768</td>
<td>0.757</td>
<td>0.967</td>
<td>0.905</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Table 3: Score correlation for individual indicators over time for 51 economies.

<table>
<thead>
<tr>
<th></th>
<th>WEF-GCI</th>
<th></th>
<th></th>
<th></th>
<th>IMD-WCY</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Eco.</td>
<td>2007</td>
<td>2008</td>
<td>2009</td>
<td>2010</td>
<td>2011</td>
<td>Eco.</td>
<td>2007</td>
<td>2008</td>
<td>2009</td>
</tr>
<tr>
<td>2007</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2007</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>0.99</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td>2008</td>
<td>0.98</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td>0.98</td>
<td>0.99</td>
<td>1.0</td>
<td></td>
<td></td>
<td>2009</td>
<td>0.94</td>
<td>0.97</td>
<td>1.0</td>
</tr>
<tr>
<td>2010</td>
<td>0.97</td>
<td>0.98</td>
<td>0.99</td>
<td>1.0</td>
<td></td>
<td>2010</td>
<td>0.92</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>2011</td>
<td>0.97</td>
<td>0.97</td>
<td>0.98</td>
<td>0.99</td>
<td>1.0</td>
<td>2011</td>
<td>0.91</td>
<td>0.93</td>
<td>0.95</td>
</tr>
</tbody>
</table>

To measure and forecast the progress for KBEs, the data can be treated as cross sectional, where many countries progress are observed at one specific point in time. Data can also be treated as a time-series data, where a specific economy 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 KBE 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, score correlation analysis are conducted for each of the five selected indicators across the years. The individual score correlation results for GCI and WCY indicators are presented in Table 3. Correlation scores for all indicators range between 91% and 99%, asserting an important fact, that KBE progress and capability does not take a sudden or sharp peak, nor a sudden nose dive, which signifies a deep infrastructural development. Therefore substantial change in the ranking between economies does not happen in short term period. It is essential to pre-process the data and use the right KBE forecasting model, to produce a high degree of prediction accuracy to capture the slight change of a KBE progress over the available short period of time.
3.2. Normalisation

Normalisations usually is used to transform or standardise different measurement units into the same unit, so they can form a comparable elements, and to avoid problems in mixing measurement units such as money, talent, birth, etc. [9]. The challenge here is to unify the score ranges between selected indicators, hence Max-Min normalisation can be applied by taking all the different scores ranges collected in the data set and transforming it to a value between 0.0 to 10.0, or -1 to 1. This normalised procedure can be achieved by using the Max-Min normalisation formula, which can be expressed as follows:

$$x'_i = \frac{x_i - minA}{maxA - minA} (new\_maxA - new\_minA) + new\_minA$$

where $x'_i$ is the normalised score for the $i^{th}$ economy, $x_i$ is the actual score, $minA$, $maxA$ are the minimum and maximum values of the scores range within index $A$; Max-Min normalisation maps the actual score $x_i$ of index $A$ to $x'_i$ in the range of the $new\_minA$, $new\_maxA$.

4. Forecasting Methodology

This study considers the competitiveness predictability in the progress of KBE between nations over time. Therefore, collected data can be treated 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 shy from the full economic prediction cycle threshold which usually requires data for more than seven years [19]. There are many individual countries to track the slight change of progress over time, hence, it is suggested to investigate three popular forecasting techniques using linear regression analysis, panel data analysis and ANNs forecasting techniques.

4.1. Linear Multiple Regression

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. Any of the five selected indicators can be regressed against itself for the previous years (time-lagged). For example the $IMD_{i,t}$ index for the $i^{th}$ economy at time $t$ can be expressed as a non-linear function $f$ below:

$$IMD_{i,t} = f(IMD_{i,t-1}, IMD_{i,t-2}, \ldots, IMD_{i,t-n})$$ (2)

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

$$IMD_{i,t} = \alpha + \beta_1 IMD_{i,t-1} + \beta_2 IMD_{i,t-2} \ldots \beta_n IMD_{i,t-n} + \varepsilon$$ (3)

where $\alpha$ 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 $\varepsilon$ 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:

$$IMD_{i,t} = f(ITU_{i,t}, WB_{i,t}, WEF_{i,t}, INS_{i,t})$$ (4)

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

$$IMD_{i,t} = \alpha + \beta_1 ITU_{i,t} + \beta_2 WB_{i,t} + \beta_3 WEF_{i,t} + \beta_4 INS_{i,t} + \varepsilon$$ (5)

$IMD_{i,t}$ is expressed as a linear function of other indices for the $i^{th}$ 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 to represent different responses and predictors between the five indices for a $t$ time period.
4.2. Panel Data Analysis

Panel data analysis or also known as Time Series Cross Sectional (TSCS) methods is considered to be the flagship of advance econometrics models. It is a technique of studying a specific entity over a particular time frame. With repeated observations of enough cross-sections, panel analysis permits to study the dynamics of change with short time series [28]. A panel data set usually consists of two parts: a cross sections such as countries, economies or individuals, and a time series, with data gathered on the same individuals, economies 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. Time Series Cross Section data consist of comparable time-series data obtained on a variety of entities. In general a TSCS regression can be expressed as follows:

\[ Y_{i,t} = \alpha_{i,k} + \beta' X_{i,t,k} + \epsilon_{i,t} \]  

where \( Y_{i,t} \) is the dependant variable (to be predicted), \( X_{i,t,k} \) is the independent variable, \( \beta' \) is \( K \times 1 \) coefficient vector and \( X_{i,t,k} \) 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): \( i = 1, \ldots, N; t = 1, \ldots, T; k = 1, \ldots, K \)

4.2.1. Panel Data Structures and types

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 [21]. 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 if the scores for each economy is spotted every year, it is called a “balanced panel” [4]. Table 4 shows a portion of the collected data to
Table 4: Balanced and unbalanced sample panel data set.

<table>
<thead>
<tr>
<th>Eco.</th>
<th>Year</th>
<th>ITU</th>
<th>WB</th>
<th>WEF</th>
<th>IMD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2007</td>
<td>4.3</td>
<td>4.0</td>
<td>1.2</td>
<td>1.8</td>
</tr>
<tr>
<td>1</td>
<td>2008</td>
<td>4.4</td>
<td>3.7</td>
<td>1.4</td>
<td>1.3</td>
</tr>
<tr>
<td>1</td>
<td>2009</td>
<td>4.3</td>
<td>3.9</td>
<td>2.0</td>
<td>0.7</td>
</tr>
<tr>
<td>2</td>
<td>2007</td>
<td>8.4</td>
<td>9.4</td>
<td>7.5</td>
<td>7.4</td>
</tr>
<tr>
<td>2</td>
<td>2008</td>
<td>8.7</td>
<td>9.2</td>
<td>7.5</td>
<td>N/A</td>
</tr>
<tr>
<td>2</td>
<td>2009</td>
<td>8.4</td>
<td>9.1</td>
<td>7.9</td>
<td>8.2</td>
</tr>
</tbody>
</table>

illustrate the difference between the balanced and unbalanced panel. The sample of the two data sets with a two dimensional panel structure are shown. In the left hand side of the table two economies/countries (1 and 2) are observed over three years for year 2007 to 2009. Since 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. In the unbalanced panel data, economy 1 was not observed in year 2009 and economy 2 score is not reported by WEF index in 2008, etc. For the estimation techniques used in this study, the distinction between balanced and unbalanced panels is relaxed.

Data source mentioned in Section 3 fits the descriptions of a panel data set, therefore, this study also suggests using the TSCS regression. This will help to symbolise the capture of relation between the different entities “economies” and to take advantage of the behaviour of these knowledge economies over time.

There are several models for the analysis of panel data including constant coefficients model also known as pooled model, fixed effect, random effect, dynamic panel, covariance structure model etc. After careful examination for each and every one of these models, the selection was laid between two major models which could suit the collected data set for this study to produce a meaningful analysis. Selected models are:

- **Fixed Effects Model** - This model is called the least square with dummy variables (LSDV) and can be represented as follows:

  \[ Y_{i,t} = \alpha_i + \beta_i X_{i,t} + \cdots + \varepsilon_{i,t} \]  

(7)
where $i = 1, 2, \ldots, N; t = 1, 2, \ldots, T$; the slopes $\alpha_i$'s are constant and they 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** - The random effects model is similar to the fixed effects model. However, the slopes $\alpha_i$'s are treated as random variables rather than fixed constants. This model is also known as the variance components model (VCM).

Whether to treat the effects of $\alpha_i$ as fixed or random number are 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 [12]. This can be hypothesised as follows:

- **H0**: No correlation between the random effects and the regressors. Accepting this allow 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 index as a dependant variable is presented in Table 5. The differences between the two sets of estimates are tested as a block using $\chi^2$ or Chi-square. The $\chi^2$ 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.
Table 5: Hausman test for IMD index: fixed vs. random effects.

<table>
<thead>
<tr>
<th>Test Summary</th>
<th>Chi-Sq. Statistic</th>
<th>Chi-Sq. d.f.</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-section random</td>
<td>105.8656</td>
<td>4</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Cross-section random effects test comparisons:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Fixed</th>
<th>Random</th>
<th>Var(Diff.)</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITU</td>
<td>1.47068</td>
<td>0.36986</td>
<td>0.00019</td>
<td>0.0000</td>
</tr>
<tr>
<td>WB</td>
<td>-10.29712</td>
<td>-1.37133</td>
<td>0.01278</td>
<td>0.0000</td>
</tr>
<tr>
<td>WEF</td>
<td>1.03969</td>
<td>0.68341</td>
<td>0.00016</td>
<td>0.0000</td>
</tr>
<tr>
<td>INS</td>
<td>-0.03584</td>
<td>-0.00603</td>
<td>0.000001</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

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 [12].

4.3. 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 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 [3].

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 [27]. 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 [18].

Training feed-forward neural network usually happens in three steps: each input $X_1, \ldots, X_n$ 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 $1, \ldots, h$ 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 [8].

SOM on the other hand is a suitable technique for multidimensional data clustering [13] and it helps to avoid the curse of dimensionality and the heterogeneity bias which arise among cross-sectional units [12]. 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 [29]. 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 [10], however, few issues need to be tackled to get efficient results. It is suggested by Thang et al. [24] 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. Collected data from 51 economies were used as inputs to the feed-
forward neural network to predict the future values for five indices. This simple mode of
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 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 [1]. The network was
trained using the Levenberg Marquardt backpropagation algorithm, which is known as one
of the most efficient ANN training algorithm [17] [2]. Figure 2 shows used data availability
and division for ANN forecasting.

4.4. **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 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 tansig transfer function and linear output neurons are 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 five 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 learn 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 distribution 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,
thus, different values were experimented till were 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. In Figure 4, each neuron is represented by a grey dot at the location of its two
weights. In this figure, the forecasted scores before and after training the SOM neurons are
shown. The figure in the left hand side shows the map before training when the map gives
all neuron equal weights, while the right hand side figure shows SOM layer of neurons after
training and how it started to self-organize, so that each neuron organises a different groups
of the input space, and related neurons get connected to the nearby neighbouring group.

5. Performance, Analysis and Results

In this section some results for indicators prediction and mapping economies are pre-
sented. Accuracy and performance of the results are measured using measured defined
below.
5.1. 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:

\[
MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{x_t - \hat{x}_t}{x_t} \right|
\]  

while

\[
MSE = \frac{1}{n} \sum_{t=1}^{n} (x_t - \hat{x}_t)^2
\]

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

5.2. The Forecasting Performance Results

After running all experiments using all methods mentioned earlier across and between different indicators, a large amounts of results combinations has been generated. Table 6 shows the overall MAPE and MSE performance results for all used techniques. Detailed training, validation and testing performance results for ANN II are presented in Table 7. By comparing the performance results one can clearly notice the difference in performance between the regressions methods and the ANN. The weakest performance obtained across all indicators was reported by the linear multiple regression (LMREG) and by contrast, the winner performances was reported by ANN II. 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).

Figure 5 show 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
Table 6: Forecast overall performance results for the selected KBE indicators.

<table>
<thead>
<tr>
<th>Method</th>
<th>WEF(GCI)</th>
<th>WB(KEI)</th>
<th>ITU(IDI)</th>
<th>IMD(WCY)</th>
<th>INS(GII)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMREG</td>
<td>8.60</td>
<td>21.93</td>
<td>13.53</td>
<td>24.74</td>
<td>33.26</td>
</tr>
<tr>
<td>TSCS</td>
<td>3.05</td>
<td>2.54</td>
<td>4.44</td>
<td>8.76</td>
<td>20.52</td>
</tr>
<tr>
<td>ANN I</td>
<td>1.86</td>
<td>1.94</td>
<td>2.10</td>
<td>3.31</td>
<td>5.32</td>
</tr>
<tr>
<td>ANN II</td>
<td>1.12</td>
<td>1.12</td>
<td>1.20</td>
<td>1.31</td>
<td>3.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>WEF(GCI)</th>
<th>WB(KEI)</th>
<th>ITU(IDI)</th>
<th>IMD(WCY)</th>
<th>INS(GII)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMREG</td>
<td>0.24</td>
<td>0.64</td>
<td>0.58</td>
<td>0.93</td>
<td>0.65</td>
</tr>
<tr>
<td>TSCS</td>
<td>0.04</td>
<td>0.04</td>
<td>0.24</td>
<td>0.15</td>
<td>0.30</td>
</tr>
<tr>
<td>ANN I</td>
<td>0.03</td>
<td>0.03</td>
<td>0.05</td>
<td>0.16</td>
<td>0.27</td>
</tr>
<tr>
<td>ANN II</td>
<td>0.02</td>
<td>0.04</td>
<td>0.05</td>
<td>0.03</td>
<td>0.24</td>
</tr>
</tbody>
</table>

Table 7: Training, validation and testing MSE performance results for ANN II model.

<table>
<thead>
<tr>
<th>ANN II Subsets Performance (MSE)</th>
<th>WEF(GCI)</th>
<th>WB(KEI)</th>
<th>ITU(IDI)</th>
<th>IMD(WCY)</th>
<th>INS(GII)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>1.131e-07</td>
<td>2.58e-04</td>
<td>9.17e-26</td>
<td>2.413e-04</td>
<td>0.110</td>
</tr>
<tr>
<td>Validation</td>
<td>0.100</td>
<td>0.074</td>
<td>0.164</td>
<td>0.059</td>
<td>0.397</td>
</tr>
<tr>
<td>Testing</td>
<td>0.0851</td>
<td>0.224</td>
<td>0.237</td>
<td>0.160</td>
<td>0.662</td>
</tr>
</tbody>
</table>

labels and the abbreviations are provided in Table 8. Figure 6 shows results obtained by ANN II for IMD, INS, WB and WEF indices.

5.3. 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 3. The coloured SOM maps presented in Figure 7(a) and Figure 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.
Figure 5: Predicted vs. real scores and one-step ahead forecasting based on the proposed ANN II for ITU index.

Table 8: Labels, abbreviations and name of economies used in this study.

<table>
<thead>
<tr>
<th>NO.</th>
<th>Eco.</th>
<th>Economy</th>
<th>NO.</th>
<th>Eco.</th>
<th>Economy</th>
<th>NO.</th>
<th>Eco.</th>
<th>Economy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ARG</td>
<td>Argentina</td>
<td>18</td>
<td>GRE</td>
<td>Greece</td>
<td>35</td>
<td>PHI</td>
<td>Philippines</td>
</tr>
<tr>
<td>2</td>
<td>AUS</td>
<td>Australia</td>
<td>19</td>
<td>HKG</td>
<td>Hong Kong</td>
<td>36</td>
<td>POL</td>
<td>Poland</td>
</tr>
<tr>
<td>3</td>
<td>AUT</td>
<td>Austria</td>
<td>20</td>
<td>HUN</td>
<td>Hungary</td>
<td>37</td>
<td>POR</td>
<td>Portugal</td>
</tr>
<tr>
<td>4</td>
<td>BEL</td>
<td>Belgium</td>
<td>21</td>
<td>IND</td>
<td>India</td>
<td>38</td>
<td>ROM</td>
<td>Romania</td>
</tr>
<tr>
<td>5</td>
<td>BRA</td>
<td>Brazil</td>
<td>22</td>
<td>INA</td>
<td>Indonesia</td>
<td>39</td>
<td>SIN</td>
<td>Singapore</td>
</tr>
<tr>
<td>6</td>
<td>BUL</td>
<td>Bulgaria</td>
<td>23</td>
<td>IRL</td>
<td>Ireland</td>
<td>40</td>
<td>SVK</td>
<td>Slovak Republic</td>
</tr>
<tr>
<td>7</td>
<td>CAN</td>
<td>Canada</td>
<td>24</td>
<td>ITA</td>
<td>Italy</td>
<td>41</td>
<td>SLO</td>
<td>Slovenia</td>
</tr>
<tr>
<td>8</td>
<td>CHI</td>
<td>Chile</td>
<td>25</td>
<td>JPN</td>
<td>Japan</td>
<td>42</td>
<td>SA</td>
<td>South Africa</td>
</tr>
<tr>
<td>9</td>
<td>CHN</td>
<td>China</td>
<td>26</td>
<td>JOR</td>
<td>Jordan</td>
<td>43</td>
<td>ESP</td>
<td>Spain</td>
</tr>
<tr>
<td>10</td>
<td>COL</td>
<td>Colombia</td>
<td>27</td>
<td>KOR</td>
<td>Korea (Rep.)</td>
<td>44</td>
<td>SUI</td>
<td>Sweden</td>
</tr>
<tr>
<td>11</td>
<td>CRC</td>
<td>Croatia</td>
<td>28</td>
<td>LTU</td>
<td>Lithuania</td>
<td>45</td>
<td>SWE</td>
<td>Switzerland</td>
</tr>
<tr>
<td>12</td>
<td>CZE</td>
<td>Czech Republic</td>
<td>29</td>
<td>LUX</td>
<td>Luxembourg</td>
<td>46</td>
<td>THA</td>
<td>Thailand</td>
</tr>
<tr>
<td>13</td>
<td>DEN</td>
<td>Denmark</td>
<td>30</td>
<td>MAS</td>
<td>Malaysia</td>
<td>47</td>
<td>TUR</td>
<td>Turkey</td>
</tr>
<tr>
<td>14</td>
<td>EST</td>
<td>Estonia</td>
<td>31</td>
<td>MEX</td>
<td>Mexico</td>
<td>48</td>
<td>UKR</td>
<td>Ukraine</td>
</tr>
<tr>
<td>15</td>
<td>FIN</td>
<td>Finland</td>
<td>32</td>
<td>NED</td>
<td>Netherlands</td>
<td>49</td>
<td>UK</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>16</td>
<td>FRA</td>
<td>France</td>
<td>33</td>
<td>NZL</td>
<td>New Zealand</td>
<td>50</td>
<td>USA</td>
<td>USA</td>
</tr>
<tr>
<td>17</td>
<td>GER</td>
<td>Germany</td>
<td>34</td>
<td>NOR</td>
<td>Norway</td>
<td>51</td>
<td>VEN</td>
<td>Venezuela</td>
</tr>
</tbody>
</table>
Figure 6: Predicted vs. real scores and one-step ahead forecasting based on the proposed ANN II for IMP, INS, WB and WEF indices.
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.
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, but 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. Also 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 proofs the volatility and the nature of these fast growing knowledge economies which draw the facts from the hype, and this is due to a low scores obtained by India and China for ICT use, skills and access, and the opposite was true for Korea, Malaysia and Argantine. 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 show is led by small economies like: Hong Kong, Singapore, Switzerland, Sweden and followed by another small neighbours economies such as Denmark, Finland, Netherlands and Norway followed by Austria, Belgium and Luxembourg; this fact should give real hope to small developing economies.

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 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. Choropleth map, usually provides a quick glimpse of how the forecasted scores are distributed across the globe. In Figure 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
Figure 8: Visualisation techniques for one-step ahead forecasted and clustered results.

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.
6. Conclusion and Future Work

Studies comparing panel data and computational intelligence techniques are sparse. This paper 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 five 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.

Giving the promising results, future study to include more economies and to compare other neural networks methods. In addition the immediate plan is to broaden the forecast to include missing data, and to use more validation measures. Eventually, the aim would be to generate a few-steps ahead unified forecast map to foresee the promising or emerging knowledge economies and to anticipate the declining ones.
References


URL http://www.millennium-project.org/millennium/SOFI.html


URL http://www.imd.org/research/publications/wcy/index.cfm


URL http://www.quantum3.co.za/CI%20Glossary.htm


28
