# Using Deep Learning Neural Networks to Predict the Knowledge Economy Index for Developing and Emerging Economies

Andres, Antonio Rodriguez and Otero, Abraham and Amavilah, Voxi Heinrich

a VSB-Technical University of Ostrava, Faculty of Economics, Department of Economics, Sokolska trida 33, 702 00 Ostrava, Czech Republic, Polytechnic School, University San Pablo CEU., Urbanizacion de Montepríncipe s/n, 28668 Boadilla del Monte, Madrid, Spain, Economics Department, Division of Social and Behavioral Sciences, Estrella Mountain College, 3000 North Dysart Road, Avondale, AZ 85392, USA

15 April 2021

# Using Deep Learning Neural Networks to Predict the Knowledge Economy Index for Developing and Emerging Economies[1]

Antonio Rodríguez Andrés- Corresponding author

VSB-Technical University of Ostrava

Faculty of Economics

Department of Economics

Sokolska trida 33, 702 00 Ostrava

Czech Republic

E-mail address: antonio.rodriguez.andres@vsb.cz

Abraham Otero

Polytechnic School, University San Pablo CEU.

Urbanización de Montepríncipe s/n

28668 Boadilla del Monte

Madrid, Spain

E-mail: abraham.otero@gmail.com

Voxi Heinrich Amavilah

Economics Department

Division of Behavioral Sciences

Estrella Mountain College

3000 North Dysart Road, Avondale, AZ 85392, USA

E-mail: vhsamavilah@gmail.com (preferred)

**April 2021**

[1]Financial support from the Czech Science Foundation (No. GA19-25280S), and the Ministry of Science, Innovation and Universities of Spain (No. RTI2018-095324-B-I00) is gratefully acknowledged. Data collection assistance was provided by Ms. Nilufar Safarova. The usual caveat applies.

**Abstract**

Missing values and the inconsistency of the measures of the knowledge economy remain vexing problems that hamper policy-making and future research in developing and emerging economies. This paper contributes to the new and evolving literature that seeks to advance better understanding of the importance of the knowledge economy for policy and further research in developing and emerging economies. In this paper we use a supervised machine deep learning neural network (DLNN) approach to predict the knowledge economy index of 71 developing and emerging economies during the 1995-2017 period. Applied in combination with a data imputation procedure based on the K-closest neighbor algorithm, DLNN is capable of handling missing data problems better than alternative methods. A 10-fold validation of the DLNN yielded low quadratic and absolute error (0,382 +- 0,065). The results are robust and efficient, and the model's predictive power is high. There is a difference in the predictive power when we disaggregate countries in all emerging economies versus emerging Central European countries. We explain this result and leave the rest to future endeavors. Overall, this research has filled in gaps due to missing data thereby allowing for effective policy strategies. At the aggregate level development agencies, including the World Bank that originated the KEI, would benefit from our approach until substitutes come along.

***Keywords***: Machine deep learning neural networks; developing economies, emerging economies, knowledge economy, knowledge economy index, World Bank

JEL: *C45; C53; O38; P41; O57*

## 1. Introduction

Measuring the knowledge economy and its underlying knowledge is a difficult task in and of itself. For developing and emerging economies, the problem is compounded by missing values as well as the inconsistency of the measures of the knowledge economy. This is huge problem, because knowledge, including technology, is central to long-run economic growth and development (Romer, 1990; 1994; Nobel Prize Committee, 2018; de la Paz-Marín, Gutierrez, and Hervás-Martínez, 2015; de la Paz-Marín, Campoy-Muñoz, and Hervás-Martínez, 2012). Knowledge is a result of past innovations (Dodgson and Gann, 2018). From Arrow (1969) it is also obvious that some innovations require purposeful investment (they are endogenous), while others are incidental to (they are positive externalities of) other economic activities and luck (they are exogenous). However, previous research also shows that geography, institutions, governance, and market imperfections, for example, contribute to the uneven distribution and stickiness of knowledge (Jaffe, Trajtenberg and Henderson, 1993; Hidalgo, 2018; Breshi and Lissoni, 2004; Singh, 2004; Hidalgo, Klinger, Barabasi, and Hausmann, 2007; Neffke, Henning and Boschma, 2011). The problem stated above particularly vexing in that innovations on which knowledge depends are hard to define and even harder to measure (Sutz, 2012; Rao, 2010), although many attempts have been made to develop appropriate indexes of knowledge as illustrated in Furman, Porter, and Stern (2002), Archibugi and Coco (2004, 2005), Khayyat and Lee (2012), Griffiths and Kichul (2008), Rao (2010), and Galebo, Plekhanov, and Silve (2015), to mention but a few.

A challenge when developing any index for emerging and, especially, for developing economies is that the macroeconomic, population, education, infrastructure data on which the index relies is often not available in a consistent manner and, when it is available, it may not always be completely reliable. This is the main reason why the World Bank's KEI does not cover these countries consistently. Being able to handle missing and unreliable data, and providing in these scenarios the best estimates possible, is therefore a desired quality for an index that attempts to quantify any facet of developing and emerging economies.

In their model of innovation performance in European regions, Hajek and Henriques (2017) concluded that strong innovation performance is a good indicator of national policy-marking in relation supporting further innovation, innovation systems, knowledge itself, and hence economic growth and development, which heighten the importance of accurate measurement of innovations. However, complicating the measurement problem further is the rapid growth of, or at least aspiration for, the knowledge economy in nearly every country, where the knowledge economy refers to an economy that depends on the quantity and quality of knowledge. The World Bank Institute responded to the challenge by constructing the KEI as an element of the general knowledge index (KI) in any economy (Parenti and Prescott, 1994).[2] Even more concerning is that to this day the KEI does not cover many developing and emerging economies, or at best does not cover them consistently. Because of this inconsistency of coverage and paucity of data, policymakers are demanding improvements on and or substitutes to existing indexes. For example, Rizk, El Said, Weheba, and de Beer (2018) have reviewed the definitions and measures of innovations and found that both macro and micro "approaches do not capture the full extent of innovation that is occurring in the context of the developing [and emerging] economies" (p. 2). Macro indicators of innovations miss micro aspects of innovations and tend to be biased towards industrialized economies. Micro measures of innovation are too focused on formal innovations, and therefore understate the vitality of informal sectors as sources and sinks of innovations in developing and emerging economies. In a related paper, Hassouna (2018) examined the methodologies used to construct 16 indices of innovation, information and communications technologies, economic environment, governance, and development. The review found the indexes deficient in a variety of areas, including the following two. First, although they follow similar data structures, and preparation and manipulation, different indexes cover different numbers of countries and different indicators of innovation. Second, developing countries, especially African but also emerging economies are not fully represented on most indexes even when these countries are ultimately consumers of intended policy. Hence, in the case of Africa, Hassouna proposes an "Innovation Activity Index (IAI)" based on three pillars: Collaboration, human skill development, and knowledge governance. Thus, although we agree that there is a need for new and comprehensive measures of innovation and knowledge, we are not quite ready yet to pour the baby out of the tub with the dirty water. Our brief literature review below shows that applications of artificial Neural Networks (NNs) and Deep Learning Neural Networks (DLNNs)

[2] For further details visit: http://web.worldbank.org/archive/website01030/WEB/IMAGES/KAM_V4.PDF. cf.

in economics are scarce, but they have provided useful insights where they were used. To the best our knowledge, there are two papers closely related to ours (Al Shami, 2011; Kuhlman, et al 2017)). Both empirical studies of NNs falls within the paradigm of unsupervised learning, while the proposal in this article falls within the paradigm of supervised learning. However, we propose a novel application the DLNN approach to innovations, knowledge generally, or technology specifically to predicting KEI in developing and emerging economies.

Regarding the need, innovations and knowledge give rise to complex and interactive economic activities (Hidalgo and Hausmann, 2009; Albaik, Kaltenberg, Alsalek, and Hidalgo, 2017). With respect to useful insights, Ojanpera, Graham, and Zook (2018) mapped content production and developed a "digital knowledge economy index" using machining learning techniques and the World Bank KEI as the training tool. *Given that background, the objective of this paper is to develop a methodology for predicting the knowledge economy index (KEI henceforth) for a few developing and emerging economies for 1995-2017 years that is robust even when some of the data on which it relies is not available*. To achieve this objective, we shall use a state-of-the-art machine learning technique, deep learning neural networks (DLNN), to predict the KEI index for developing and emerging countries using a set of economic variables based on the World Bank Knowledge Assessment Methodology (KAM) (Chen and Dahlman, 2006). The prediction of the KEI index from economic data has been studied for developed countries, where it can be assumed that all the economic variables used in the prediction are available (Al Shami, 2015). However, data from emerging and developing countries contains a large share of missing values. In the specific case that concerns us, of the 989 country-year combinations that will be used in this study, only 57 of them present complete values. Furthermore, multiple missing values are typically present in most of the country-year combinations. To overcome this situation, a data imputation strategy based on the K closest neighbors algorithm was used.

The remainder of the paper is organized as follows: In the next section, we briefly review the relevant literature. First, we outline the basic structure and uses of the NNs (Subsection 2.1). Secondly, we describe the deep learning neural networks – DLNNs – (Subsection 2.2). Finally, in Subsection 2.3 we indicate how NNs and DLNNs apply to economics and related areas. Section 3 presents our research methodology, including the economic variables and data, the specific methods we use, and how we train and validate the DLNN. Section 4 presents and discusses the results, while the final section (Section 5) summarizes the main conclusions.

## 2. Review of Relevant NN and DLNN Literature

Machine learning is a field of computer science that focuses on the development of computational techniques capable of finding regularities in (typically) large datasets, and on how to exploit these regularities to take decisions about new data. Machine learning techniques can be divided into two main types: unsupervised learning and supervised learning (Bishop, 2006). In a supervised learning problem, there is a well-defined analytical goal (for example, predicting the KEI for those country-year combinations for which it is not available), and the computational technique requires a labeled dataset to be able to learn; i.e., a dataset for which the correct answer to the problem is known (for example, the KEI values for those country-year combinations for which it is available). Unsupervised learning tries to learn unknown regularities in a dataset without needing a well-defined analytical goal or a labeled dataset. An example of unsupervised learning could be finding sets of countries whose economic evolution has been similar in the last decade. Initially, we do not know how many groups of countries there may be (those groups may not even exist) or what are the specific causes of them having a similar evolution. While some machine learning technique can be used only in the context of supervised learning, or only in the context of unsupervised learning, NN and DLNN can be used in both scenarios. However, in the rest of this section, whenever we refer to machine learning, we shall do it in the context of supervised learning, since that is the kind of problem that is being addressed this paper.

### 2.1 Introduction to Neural Networks

NNs is a machine learning technique loosely inspired by the functioning of animal brains. A natural neuron is a cell that has a series of input connections (dendrites) connected to the cell body, and a single exit termination (typically much longer than the dendrites) called the axon (Ashwell, 2012; Zhu, 2017). The dendrites of a neuron are typically connected to axons of other neurons, and receive stimuli from them. The body of the neuron integrates all these stimuli and based on them may or may not fire an action potential, i.e., it generates an electrical activity that is transmitted along its axon, which is usually connected to the dendrites of other neurons. A stimulus that reaches a given dendrite can have either an excitatory function (contributes to the receptor neuron firing) or inhibitory (contributes to the receptor neuron not firing). The strength with which a given stimulus excites or inhibits another neuron can change with time, and these changes enable animal brains to learn new things (Ashwell, 2012). An isolated natural neuron is a relatively simple system: it receives some inputs, it integrates them, and it decides whether to fire an action potential or not based on them. But, when many of these neurons connect with each other they can do remarkable things such as recognizing words on a paper or people's faces, talking, writing a poem, producing a work of art, or carrying out a scientific discovery.

In a similar way, artificial NNs have a series of inputs $x_i$ ("dendrites"), each of which is multiplied by a weight $w_i$ (Zurada, 1992). The module of $w_i x_i$ represents the magnitude of the influence of that input over the artificial neuron, and its sign represents an excitation (positive sing, a contribution to fire) or inhibition (negative sing, a contribution not to fire). The body of the artificial neuron adds the n inputs it receives, multiplied each by its corresponding weight ($\sum_{i=1}^{n} w_i x_i$), adds some additional value $w_0$ called bias, and then applies a non-linear function over this sum, $f(w_0 + \sum_{i=1}^{n} w_i x_i)$, to determine the output (the "axon") of the neuron. This non-linear function is key to the functioning of the NN since it permits representing non-linear behaviour in the data. If the function were linear, the NN would be equivalent to a linear regression. Gershenson's (2003) guide to ANNs is an excellent introductory text for beginners. Mathematically, a single artificial neuron is underwhelmingly simple, and as a machine learning tool, it is a poor tool. But when many of these neurons are connected in a net in such a way that the inputs of a neuron are the outputs of a layer of previous of neurons, a system emerges that has a surprising potential for learning (Hornik, 1989). NN can be used to learn unknown regularities in a data set (unsupervised learning), or to learn to predict some data, being necessary in this case a dataset for which the value to predict is previously known (supervised learning) (Bishop, 2006).

McCulloch and Pitts (1943; Hebb, 1949) were the first to develop a computational model of a network of neurons using threshold logic. In the 50s and 60s there were many theoretical advances regarding NNs (Hush, 1993), but it was in 1975 when Paul J. Werbos proposed the backpropagation algorithm to train NNs in his PhD thesis (Werbos 1990; Rumelhart, Hinton, and Williams, 1986; Hecht-Nielsen, 1987; Hinton, 1987; Hopfield and Tank, 1986; Hopfield, 1982). This algorithm enables the efficient training of a NN from a set of examples, i.e., data for which the desired output for the NN is known. It is an iterative and recursive algorithm that calculates updates to the weights of the connections between neurons ($w_i$) using a gradient descendant method which tries to minimize the distance between the output produced by the NN, and the desired output. Although this algorithm was a significant advance, it suffers the vanishing gradient problem (Hochreiter, 1998). The backpropagation algorithm uses partial derivatives updating the network weights, and under certain circumstances these derivatives can be vanishingly small, effectively preventing the weights from changing, and therefore preventing the network from learning (Hochreiter, 1998). Furthermore, even when this problem does not hamper the training of the network, the backpropagation algorithm can sometimes overfit the network (Tzafestas, 1996). It is said that a NN has overfitted when it "memorizes" the training dataset and is able to correctly predict the desired value for it, but it makes poor predictions when presented with different data from the training set. Overfitting is more likely the more complex the topology of the network is. Both problems cause that, in practice, the backpropagation algorithm cannot train complex network topologies, limiting the learning capacity of the NN.

## 2.2. Introduction to Deep Learning Neural Networks

In the 2000s, and especially in the present decade, a series of theoretical and computational advances have been made that have enabled the training of NNs with significantly more complex topologies than what the backpropagation algorithm enables. These advances include new non-linear activation functions that are less likely to suffer from the vanishing gradient problem (Clevert, 2015), new regularization techniques such as dropout, batch normalization, and data-augmentation that make NN less prone to overfitting (Zaremba, 2014); new stochastic gradient descent techniques that allow for more robust optimization of the network weights (Johnson, 2013); and the increase in computing power of computers which makes feasible the use of more complex training algorithms and the processing of larger data sets. All these advances have enabled the training of NNs of significantly more complex network topologies and larger sizes than those allowed by the backpropagation algorithm, leading to the emergence of the field of Deep Learning Neural Networks (DLNN). A DLNN is a NN that has at least two hidden layers of neurons. A NN always has an input layer of neurons that receives the input data, and an output layer responsible for outputting the prediction. The layers of neurons that lie in between are called hidden layers. DLNN has produced a revolution in the field of machine learning due to its great capacity for generalization and learning. These techniques have dramatically improved the performance of state-of-the-art algorithms fields as diverse as image classification, speech recognition, drug discovery and genomics (LeCun, 2015).

### 2.3. Applications of Neural Networks to Economics and Finance

Neural networks (NNs) have been used in marketing, retail and sales, and banking and finance for many years now. White (1988) utilized neural networks to predict IBM daily stock returns. Since then the performance of NNs in forecasting has been compared to that of economic and econometric models (Kuan and Liu, 1995; Moshiri and Cameron, 2000; Hajek, 2017; Moghaddan, et al 2016). Nakamura (2004) forecasted US inflation and concluded that "neural networks outperform univariate autoregressive model on average for short horizons of one and two quarters" (p. 373; cf McNelis and McAdam, 2005). NNs for economic prediction are also not new research activities as Herbrich, Keilbach, Graepel, Bollmann-Sdorra and Obermayer's (1998) overview clearly shows (cf. Verkooijen, 1996). Hajek, Olej and Myskova (2013) applied several NNs and support vector regressions (SVRs) to "predict the yearly change in the stock price of U.S. firms," finding that NNs and SVRs perform better than linear regression models. In a recent article Chatzis, Siakoulos, Petropoulos, Stavroulakis, and Vlachogiannakis (2018) employed deep and statistical machine learning techniques to forecast episodic stock crises. They found robust results and concluded that deep learning neural networks produce accurate classifications and policy tools. Liu (2019) has demonstrated further the usefulness of deep learning neural network techniques for financial decisions in forecasting volatility that is essential to optimizing risky portfolios.

The value of machine learning methods, especially at the intersection of theoretic econometric models and applied algorithmic approaches is growing fast (Athey and Imbens, (2019; Athey, 2017; 2019; Varian, 2014). Nonetheless, ANNs applications in economics are still less frequent. Among the few existing studies, Swanson and White (1997) have suggested model selecting criteria for forecasting with artificial NNs. Another attempt has been made by Mihaylova (2018) who describes how ANNs are applied in economics and finance for prediction, classification, and modeling purposes (Nair, 2014; Kline, 2019; Zanak and Becerra-Fernandez, 2005). In fact, there are only two articles closely related to ours. Al Shami (2011) designed a unified knowledge economy competitiveness index that reflects the overall rate of knowledge in an economy using a NN which helped to identify the data that made significant contributions to their knowledge indicator. Note that this use of NNs falls within the paradigm of unsupervised learning, while the proposal in this article falls within the paradigm of supervised learning. The closet to our approach in this vein of research is Kuhlman, et al 2017) who used a Group Lasso model to predict the "levers" of innovation in 150 countries. While we cover the same general area, they use unsupervised, we deploy supervised, learning methods. From the literature reviewed we are unaware of DLNN applications to innovations, knowledge generally, or technology specifically, and most certainly not to predicting KEI in developing and emerging economies. This is a significant gap in the literature and one reason why the current paper participates in ongoing efforts to fill the gap.

## 3. Research Methodology

### 3.1. Measuring knowledge economy

Following de la Paz-Marín, Gutierrez, and Hervás-Martínez (2015), in the current study, we are taking as a benchmark the KEI index and the choice of variables underlying it that are used by the World Bank. This index measures the knowledge economy at country level and is one of the most common composite indicators assessing the knowledge economy. At the present, the World Bank (2009) sets the 109 structural and qualitative variables for 146 countries in the world to quantify their level of knowledge economy over time. For that purpose, they employ the KAM methodology that proposes a set of proxy variables organized into four main pillars to describe the knowledge economy of a country:

1. **An economic and institutional regime** providing appropriate incentives for the efficient use of new knowledge and the flourishing of entrepreneurship. In our current study, three variables represent this pillar: rule of law, regulatory quality, and tariff and non-tariff barriers.
2. **An educated and skilled population** to share, generate, and use of knowledge as well. Education is proxied by secondary, tertiary, and primary enrollment rates.
3. **An information and communication infrastructure** to facilitate and disseminate the effective creation, and information. This dimension is proxied by variables such as: internet users per 1000 people, broadband internet subscribers, and number of telephone lines per capita.
4. **An efficient innovation system** of research centers, firms, universities, and other public bodies to promote knowledge creation and knowledge diffusion. Patent applications and scientific journal articles are employed as proxy variables for measuring this dimension.

In this paper we adopt a more holistic approach of the wide spectrum of factors relevant to a knowledge economy and categorizes them in the four pillars above mentioned that is in line with Parcero and Ryan (2017). According to the data published by the World Bank (World Bank, 2012), European countries generally dominate in ranking of knowledge economies. One of the problems is that this index is non-available on continuous basis. We can only gather statistical information for separate years. Moreover, most of the missing values correspond to Sub-Saharan African countries. The latest year available is for 2012.

### 3.2. Economic Variables and Data

Table 1 provides the definitions and source of the variables. Again, the choice of variables was suggested by the KAM and we have no reason to dispute its appropriateness. It is quite possible, perhaps even likely, that inclusion of other variables would question the robustness of our approach one way or the other. However, that is a different problem; our objective as stated previously is to demonstrate that where data is missing the DLNN technique can predict KEI, and other economic variables like, to the benefit of both policy and further research. The choice variables suggested by the KAM also guided our use of the nearest neighbor method for missing value imputation.

Table 2 displays the means for the variables employed in the empirical analysis. We use available data on these variables. However, preliminary analysis shows many missing values. For example, in the dataset there are a total of 3,286 missing values (30.2% missing values). Of the 989 country-year combinations, only for 57 of them all the data is available. Table 3 shows the distribution of missing values across variables; the number of patents in the country (PATENT, 76.7%), broadband penetration (FIXBI, 45.5%), education at the university level (TERTIARY, 42.5%), the number of publications in scientific journals (STJOU, 39.6%) are the variables that are most often missing. In the case of FIXBI, it should be mentioned that such a high value of missing data is misleading; the study period begins in 1995; at that time, the broadband Internet was practically non-existent in the whole world. Up to the year 2000, none of the countries studied reported this data, and until 2005 most of the countries are not reporting this data. If we only count FIXBI missing values once a country has reported this data for the first time, it is only missing in 2.1% of the time. This situation clearly requires a robust procedure to handle missing values, and this paper develops a way of predicting the KEI for a number developing and emerging economies using a state-of-the-art DLNN in combination with a data imputation procedure based on the K-nearest neighbor algorithm. The DLNN is a reasonable method to use because it allows us to "train" it to "deep-learn" from existing KEI data, generate the missing data, and "discover knowledge" that would enable us to predict KEI.

**Table 1 around here**

**Table 2 around here**

**Table 3 around here**

## 3.3.Specific Methods

### 3.3.1 Imputation of the missing values

Missing values hamper effective policy and research. For this reason, traditionally economists have resorted to dummy variable and switching regressions (Kennedy, 1992; Lim, Narrisetty, and Cheon, 2017), empirical functions like bootstrapping (Efron, 1994), and cointegration approaches (Harris, 1995) to circumvent the missing value trap. As McDonough and Millimet (2016) show these methods are lacking, and in our judgment the DLNN offer some advantage. The first step to try to prepare the data for the application the DLNN is to impute the 30.2% of missing values. A common strategy is to use the mean of the attribute as the imputation value (Honaker and King, 2010; Kleinberg, et al, 2018; Mullainathan and Spiess, 2017). However, mean imputation for a non-stationary data, i.e., for data that evolves over time, is a suboptimal solution. For example, in the problem that concerns us, the state of the telecommunication infrastructures of the different countries studied (and of the world in general) for the years 1995 and 2017differs significantly. Using the mean of the indicators related to telecommunication infrastructures would tend to overestimate their state development at the beginning of this period, and to underestimate it at the end, thus distorting the data, and compromising the quality of the subsequent analysis. Even if we decide to calculate the mean within a temporal interval (for example, every year), *does it really seem like a good idea to estimate, for example, the number of patents produced in Tanzania in 2014 (one of the missing values) by calculating a mean value that includes the patents produced in 2014 in countries such as China, Brazil, Russia, or the United Arab Emirate?* The answer is clearly, “No.” Therefore, if our goal is to estimate the number of patents in Tanzanian in 2014 it would make better sense to the use number of patents produced in 2014 in countries like Tanzania in the rest of the indicators, such as for example Kenya, Uganda, or Zambia, and also the number of patents in Tanzania in 2015 for which data is available.

This is precisely the idea of value imputation by the k nearest neighbors’ algorithm (Zhang, 2012). Applied to our problem to impute a missing value in one of the indicators of a given country, first the algorithm looks for the k countries that are more like the one with the missing value, and then it calculates the mean value for that indicator among those k countries. To find the k-most similar countries, first each of the indicators is standardized to make their values comparable. Then the Euclidean distance between the standardized vector of the country with the missing value and the vector representing all other countries for which the value is available is calculated. Then the k countries with the lower Euclidean distance are selected, and the mean value of the indicator for the k countries is the value of the imputation. In the vector representing each country the value of the year is included, thus not only countries like the tone with the missing value will tend to be selected for the imputation, but also data which is temporally close to the missing value (including data from the same country in adjacent years, if available).

For the imputation, the fixed broadband (FIXBI) parameter, which is missing in 45.5% of cases, deserves special attention. As we have already explained, the reason why there are so many missing values for this parameter is that in the initial years of this study broadband was almost non-existent throughout the world. For the first five years of the period under study no country reports the FIXBI data. Thus, for the first years there will be no other temporally close value of FIXBI to be used for imputation with k nearest neighbors. Thus, it does not seem a good idea to calculate the mean of broadband penetration of several countries in 2005 to impute a value of 1995. Therefore, and under the hypothesis that typically when the countries do not report this data is because they still do not have broadband (once a country starts reporting FIXBI, it is missing only in 2.1% of the cases) to impute this parameter we proceeded as follows. A value of 0 was set for all countries for all years starting in 1995 until the first year it reports a value for FIXBI. We verified that the first reported value is extremely low (almost always below 0.1), which is consistent with the previous values being 0 or very close to 0. The remaining 2.1% of FIXBI missing values were imputed through k nearest neighbors, in a similar way as the missing values from the rest of the indicators.

Prior to the imputation, the data was standardized; the ranges of the original characteristics varied enormously; for 2017, example the maximum value for the indicator of the patents is 0.021, while the maximum value for TELEP is 326.5. Then the k-nearest neighbours imputation with k = 5 was carried out using the R package caret (Kuhn, 2015). To facilitate the subsequent interpretation of the data, once the imputation was made, standardization was reversed to recover the original values.[3]

### 3.3.2 Training and validation of the deep learning neural network

Amazon's H2O library (Cook, 2016) was used for the construction of a DLNN with three hidden layers with 30, 50 and 30 neurons, respectively. The non-linear activation function chosen for the neurons was the exponential rectifier linear unit function due to its advantages over more traditional functions such as Relus and ELSs. This unit function tends to have a faster convergence and better generalization capability than the alternatives (Clevert, 2016). The gradient descendent method ADELTA was used (Zeiler, 2012) for network parameter optimization. This algorithm has the advantages over standard stochastic gradient descendent of providing a per-dimension learning rate with a computational cost only slightly higher. In the ADELTA algorithm the decay of the previous parameter updates was set to 0.995, and the momentum parameter to 1.0E-8. Lasso regularization was used to prevent overtraining (Tibshirani, 2015). In Lasso regularization, the

[3] RStudio script in available on the Internet as a csv file with the imputed data for other researchers who want to use them.

objective function that gradient descendent tries to optimize is the quadratic error between the output of the DLNN and the expected output plus the absolute value of the magnitudes of all the weights of the DLNN multiplied by a parameter L1. This last term is the penalty introduced by Lasso regularization, and it tends to shrink towards to zero the weights of the DLNN associated with the less important variables, carrying out a kind of feature selection (if a weight is zero the corresponding variable has no impact on the subsequent neuron). Lasso regularization is an effective technique to prevent overtraining of the network. In our DLNN we set the parameter L1=0.001. The loss function was the quadratic error between the output produced by the network and the desired value, which in our case is the known KEI for a country each year. Two hundred and fifty (250) epochs were used for training; i.e., during the DLNN training the data was presented 250 times to the gradient descendent algorithm.

Each one of the feature vectors presented to the DLNN was be made up by the country, the year, and the 12 indicators that are being considered in this study. All these variables are metric except the country. The country was coded by a binary variable that takes the value 1 when the vector corresponds to that country, and 0 for all other countries. In our study, we have a total of 71 developing and emerging countries, thus 71 binary variables will be needed to code the country. Therefore, the total number of inputs to the DLNN will be 84: 71 to code the country, 1 for the year, and the 12 indicators considered. The output will be a single neuron providing the value of the predicted KEI. There is no well-defined approach to find the optimal number of intermediate layers, and of neurons in each intermediate layer (Karsoliya 2012; Albawi 2017). In our case, given the relatively low number of instances available for training, and after testing several options, we opted for a relatively simple architecture for the hidden layers: three hidden layers with 60, 30 and 60 neurons. Therefore, the complete topology of the neural network will consist of five fully connected layers with 84, 30, 60, 30 and 1 neurons in them, respectively. The 84 entries of the DLNN and the term of bias are analogous to a linear regression with 84 explanatory variables (inputs); while the output neuron would be equivalent to the dependent variable of the regression (the KEI in our case). However, a particularly important distinction is that each neuron of the DLNN applies a non-linear function on the linear combination of weights multiplied by the variables that it receives. Having several consecutive layers applying nonlinear functions allows the DLNN to capture highly non-linear regularities in the data. To validate the performance of the network, a 10-fold validation was used, utilizing the 212 country-year combinations for which the World Bank KEI is available.

## 4. Results and Discussion

The 10-fold validation of the DLNN yielded a quadratic error of 0.382 +- 0.065 and an absolute error of 0.299 +- 0.044.[4] Figure 1 represents the KEI predicted by the DLNN versus the values available from the World Bank for all countries, (Pearson correlation coefficient $\rho=0.994$, p-value <1.0e-10 for all countries; $\rho=0.993$, p-value <1.0e-10 for developing economies; $\rho=0.991$, p-value= 1.0e-10 for emerging economies; and for Central European countries, $\rho=0.900$, p-value = 1.96e-09). Moreover, both error types reveal that the results are robust and of high/reasonable quality. This means that the predictive power of DLNN is exceptionally good. For countries that have data the distance between actual KEI values and KEI predicted values is exceedingly small irrespective of geographical region setting. Consider the following examples in 2012. Angola's actual KEI was 1.1 and the DLNN method predicted it at 1.3. The difference is only one-tenth of a percent for Algeria, and 5.43 (actual) versus 5.51(predicted) for Argentina. Whereas DLNN over-predicted KEI in China and Russia by 0.14 and 0.23, respectively, it under-predicted for Saudi Arabia (-0.27). KEI is also understated for the Czech Republic, Poland, and Hungary, but this result is systematic to the institutional framework of Central European countries.

Generally, Figure 1 shows that DLNN predicts KEI very well for all countries for and developing countries in the sample very well. DLNN prediction for emerging economies is also reasonable, but these economies divide into four country-year groups in the figure. In that case the predictive power of DLNN is weaker the smaller the number of K closest neighbors, which implies that the number of K constrains deep learning and thereby reduces DLNN predictive power. Although we do not have objective metrics for the accuracy of the imputation made for the missing values, the quality of the KEI prediction made by the DLNN depends on how close those imputed values are to the true values. If the imputations significantly distort the feature vectors, it would not be possible to adequately train the DLNN, nor would it be possible to make accurate predictions using wrongly imputed feature vectors, which therefore do not adequately represent the country's situation. The quality of the imputation in our approximation seems to depend on how similar the K closest neighbours chosen for this process are. In the case of developing and emerging countries, because there are a greater number of countries, the K closest countries tend to be more like the country that needs imputation. In the case of the Central European countries, given that there are fewer countries in that region, when choosing the K closest neighbours, it is more likely that it will be necessary to choose some countries that, despite being among the K most similar countries, they are not really that like the imputed country. This could explain the worsening in performance of the DLNN in predicting the KEI of the Central European countries (see Figure 1; Pearson correlation coefficient for developing economies; $\rho=0.991$ for emerging economies; and $\rho=0.900$ for Central European countries). It is possible that our approximation would have improved by using a lower K value for the imputation of the Central European countries. This suggests that our approach may benefit from using an adaptive K value depending on the density of instances in the region of the feature space where the country to be imputed is located. If the density of the region

[4]The 32-page file containing all the predictions made for the KEI as well as variable definitions are available online, and or by email upon request.

is high, the use of higher K would be favoured since there are many similar neighbours, but if it is low, lower K values would be used to avoid choosing neighbours that are not really that similar.

Another key finding is that the DLNN shows that KEI has grown for most countries in this sample during the 1995-2000 but slowed post 2000. The slowdown appears steepest for low-income countries most of them in Sub-Saharan Africa than in high-income emerging economies like the Czech Republic. Although not uniformly, KEI in emerging economies experienced a decline post-2012. This finding calls for in depth studies of KEI pillars themselves.

***[Insert Figure 1 here]***

One of the first works in the literature to use of neural networks for the prediction of economic data was Pao (2006) who concluded that neural networks had a capability equivalent or superior to that of traditional regression techniques for the prediction of time series in economics. In (Kuhlman, 2017) a group lasso predictive model was trained to generate an index developed by the authors themselves that aims to measure the level of innovation in each country. Although using machine learning to automatically generate indices that characterize the overall state of the country is an interesting application, it remains to be determined whether these indices can gain global adoption and acceptance, as happens it with indices created by recognized economic institutions, such as the KEI of the World Bank.

In Shami (2011) a system was developed that could predict a country's KEI using, among other data, three other indices that measure the country's knowledge economy (developed by the World Economic Forum, the International Institute for Management Development, and the International Telecommunication Union, respectively). While predicting one index from others is an interesting problem, it does not solve the common problem in emerging and developing countries where often many data needed to calculate any index is missing; in these contexts, calculating KEI from three other different indices may even be counterproductive, since it is more likely that data is missing for the calculation of the three indices than for the calculation of a single one. In Ahmadi (2019) a multilayer perceptron together with an adaptive neuro-fuzzy inference system and gene expression programming are used to predict the economic growth of a single country (Iran) between 1993 and 2013 using indicators related to the knowledge economy. Although the accuracy of the predictions is high, the need for training a model for each country hampers the applicability of this approach.

The closest work to ours is Shami (2015), where the KEI is predicted using neural networks (among other techniques, although neural networks are the ones that produced the best result) for 51 developed countries from 2007 to 2009 (five years) and without considering situations where missing values occur. The authors of this work pointed as a line of future work "to broaden the forecast to include missing data". Our work largely complementary to this, since it presents a methodology that allows handling missing values and focuses on the KEI prediction for emerging and developing economies, for which being able to make predictions with incomplete information it is essential. Despite having to impute the missing data, working with more countries (71 vs. 51), and working over a much longer period (22 years vs. 5), our mean quadratic error, 0.382 + - 0.065, is comparable with the one obtained by Shami (2015), 0.224.

## 5. Conclusions

The objective of this paper is to predict KEI for developing and emerging economies even in the cases where there is a high proportion of missing values. To accomplish the objective, we have used deep learning neural networks (DLNN), in combination with a data imputation procedure based on the k-Nearest Neighbors (kNN) to address the high prevalence of missing values in the economic data of these countries (94.25% of the feature vectors we worked with had missing values, most of them having multiple missing values). We found robust results; the predictions of KEI in 71 countries over the 1995-2017 period yielded a mean quadratic error of 0.382 +- 0.065 and a mean absolute error of 0.299 +- 0.044. However, there is a slight difference across countries at different levels of economic development. KEI has tended to rise for all countries pre-2000 and to fall post-2000, with the fall being steepest for low-income countries up until 2012 when it started to decline in emerging economies as well. Previous work on the role of institutions in economic performance like North (1990), Landes (1998), Blackburn and Forgues-Puccio (2010), and Tchamyou (2016) would suspect differences in institutional changes are the likely causes of the differences in KEI changes. However, why KEI has fallen for both developing and emerging economies opens a new frontier for further research.

DLNN prediction of KEI is weakest for the Central European countries. We suspect three implications for this result. First, the result might be due to fewer K closest neighbors in this group than there are for larger country groups in this sample, which constrains model training and hence opportunities for deep learning. Second, it is also likely that Central European countries just do not belong in these neighborhoods; they have more in common with industrialized Europe than with developing and emerging economies. Finally, these economies are “emerging” not in the sense of KEI, but rather in the sense that they are transitioning from the command to the market economic systems.

Our analysis also offers several contributions to literature at both the policy and research levels. In terms of policy, it filled in gaps due to missing data thereby allowing for effective policy strategies. Following the methodology, we suggest in this paper, every country can potentially derive a complete series of its KEI series. Given a complete series, at the aggregate level development agencies, including the World Bank that originated the KEI, can put our approach to use until substitutes come along. Within countries we conclude that further research in needed on the components of KEI to help understand the deep, rather than just the proximate, causes of KEI differences across countries. Such research might be the starting point with a careful examination of economic growth adjusted for the KEI differences. This is like estimating the welfare effects of knowledge (technological) divide. In the short run the marginal benefits are likely higher of low KEI economies than they are for high KEI economies.

Another important implication of this paper for both policy and future research is that if we can derive by the DLNN technique a complete series of KEI as a dependent variable, we can be then begin asking earnest questions about the determinants of technological knowledge and quantify their marginal effects. The literature we reviewed in this paper is convincing that knowledge is essential to long-run growth. However, both policy and research remain incomplete in illustrating what determines KEI itself. In developing and emerging economies, the problem has been the lack of complete KEI data and the inability of most traditional methods to deal with that problem. The methodology of this paper takes both policy and future research one step forward.

One weakness of this exercise is that we were not able to compare this method algorithmic method to theory-based methods – another opportunity for future research. A possible remedy for this weakness is to apply DLNN to predict other commonly used indices and use the results to select the best performing index among alternatives. Until all that is done, we are convinced that the DLNN may be one of the best among competition for predicting KEI under conditions of missing values.

Lastly, to avoid potential shortcomings of composite indicators to assess the level of KE across countries and over time, we propose as future extension an index that rests on the linear innovation model (input-output perspective). Thus, our proposed index would be composed of three main sub-indices: the first two, innovation enablers and innovation outcomes, capture the domestic input/output effort in innovation, whereas the third dimension captures the international diffusion of technology. Combined, R&D expenditures, and institutional, social, and structural factors become innovation enablers that, in low-income countries particularly, have to be taken into account to complement conventional metrics and to capture the role played by local capabilities in the process of creation and diffusion of knowledge. The methodology here presented can be useful as a tool for decision-making to compare results with other known innovation indices.

## References


Ahmadi, M., Jafarzadeh-Ghoushchi, S., Taghizadeh, R., and Sharifi, A. (2019). Presentation of a new hybrid approach for forecasting economic growth using artificial intelligence approaches. *Neural Computing and Applications*, 31(12), 8661-8680.

Albawi, S., Mohammed, T. A., and Al-Zawi, S. (2017, August). Understanding of a convolutional neural network. In 2017 International Conference on Engineering and Technology (ICET) (pp. 1-6).

Albeaik, S., Kaltenberg, M., Alsaleh, M., & Hidalgo, C.A. (2017). Improving the economic complexity index. Cornell University. https://arxiv.org/ftp/arxiv/papers/1707/1707.05826.pdf. Accessed May 15, 2019.

Allison, P.D. (2001). *Missing data*. Thousand Oaks (CA): Sage.

Al Shami, A., Lotfi, A., Coleman, S., and Dostál, P. (2015). Unified knowledge based economy hybrid forecasting. *Technological Forecasting and Social Change*, 91, 107-123.

Al Shami, A., Lotfi, A., Lai, E., & Coleman, S. (2011, April). Unified knowledge economy competitiveness index using fuzzy clustering model. In *2011 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFEr)* (pp. 1-6). IEEE

Archibugi, D., & Coco, A. (2004). A new indicator of technological capabilities for developed and developing countries (ArCo). *World Development*, 32(4): 629-654.

Archibugi, D., & Coco, A. (2005). Measuring technological capabilities at the country level: A survey and a menu for choice. *Research Policy*, 34(2): 175-194.

Arrow, K.J. (1969). Classificatory notes on the production and transmission of technological knowledge. *American Economic Review,* 59(2): 29-35.

Ashwell, K.W. (2012). The brain book: Development, Function, Disorder, Health (pp. 230-233). New York: Firefly Books.

Athey, S. (2017). Beyond prediction: Using big data for policy problems. *Science* 355(6324): 483–485.

Athey, S. (2019). The Impact of Machine Learning on Economics. The Economics of Artificial Intelligence: An Agenda. Ajay Agrawal, Joshua Gans, and Avi Goldfarb, editors (p. 507 -547). Conference held September 13-14, 2017.

Athey, S., & Imbens, G.W. (2019). Machine learning methods economists should know about. Available at https://arxiv.org/abs/1903.10075. Accessed June 8, 2019.

Barlow, H.B. (1989). Unsupervised learning. *Neural computation,* 1(3): 295-311.

Becker, J., & Prischmann, M. (1993). Supporting the design process with neural networks. A complex application of cooperating neural networks and its *implementation Journal of Information Science Technology*, 3 (1): 79-95

Bishop, C.M. (2006). *Pattern Recognition and Machine Learning*. Springer.

Blackburn, K., & Forgues-Puccio, G.F. (2010). Financial liberalization, bureaucratic corruption, and economic development. *Journal of International Money and Finance*, 29(7): 1321–1339.

Breschi, S., & Lissoni, F. (2004). Chapter 28 Knowledge Networks from Patent Data: Methodological Issues and Research Targets. In *Handbook of Quantitative Science and Technology Research The use of Publication and Patent statistics in Studies of S & T Systems*. Edited by Henk F. Moed, Wolfgang Glänzel, and Ulrich Schmoch. Chapter 28.

Caudill, M. (1987). Neural networks primer, Part 1. AI Expert 2 (12): 46-52.

Chalfin, A., Danieli, O., Hillis, A., Jelveh, Z., Luca, M., Ludwig, J., & and Mullainathan, S. (2016). Productivity and selection of human capital with machine learning, *American Economic Review*, 106(5): 124–127.

Chen, D.H.C., & Dahlman, C.J. (2006). The Knowledge Economy, The KAM Methodology and World Bank Operations. http://documents.worldbank.org/curated/en/695211468153873436/pdf/358670WBI0The11dge1Economy01PUBLIC1.pdf . Accessed May 15, 2019.

Clevert, D.A., Unterthiner, T., and Hochreiter, S. (2015). Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv:1511.07289.

Cook, D. (2016). *Practical Machine Learning with H2O: Powerful, Scalable Techniques for Deep Learning and AI*. O'Reilly Media, Inc.

De la Paz-Marín, M., Campoy-Muñoz, P., & Martínez, C.H. (2012). Non-linear multi-classifier model based on Artificial Intelligence to predict research and development performance in European countries, *Technological Forecasting and Social Change*, 79(9): 1731-1745.

De la Paz-Marín, M., Gutiérrez, P.A, Martínez, C.H. (2015). Classification of countries' progress toward a knowledge economy based on machine learning classification techniques. *Expert Systems with Applications*, 42(1): 562-572.

Dodgson, M., & Gann, D. (2018). *Innovation: A very short introduction*. Oxford: Oxford University Press.

Efron, B. (1994). Missing data. Imputation, and bootstrap. *Journal of the American Statistical Association*, 89(426): 463-475.

Furman, J.L., Porter, M.E., & Stern, S. (2002). The determinants of national innovative capacity. *Research Policy* 31(6): 899-933.

Gelebo, E., Plekhanov, A., and Silve, F. (2015). Determinants of frontier innovation and technology adoption: cross-country evidence. Working Paper No. 173. European Bank for Reconstruction and Development (EBRD).

Gernert, D. (1998). Classification and prediction by neural networks, and their social implications. In Computer Modeling of Social process. Edited by Win B.G. Liebrand, Andrzej Nowak and Rainer Hegselman. London: Sage Publications. Chapter 5.

Gershonson, C. (2003). Artificial neural networks for beginners. https://arxiv.org/ftp/cs/papers/0308/0308031.pdf. Accessed June 18, 2019.

Griffiths, M.D., & Kickul, J. (2008). The socioeconomic determinants of innovation: An empirical examination. *Entrepreneurship and Innovation*, 9(4): 231-240.

Hajek, P., & Henriques, R. (2017). Correction: Modelling innovation performance of European regions using multi-output neural networks. *PLOS ONE*, 12 (2): e0189746.

Hajek, P., Olej, V., & Myskova, R. (2013). Forecasting stock prices using sentiment information in annual reports – A neural network and support vector regression approach. WSEAS Transactions on Business and Economics, 10(4): 293-305.

Harris, R.I.D. (1995). *Cointegration Analysis in Econometric Modeling*. London: Harvester Wheatsheaf.

Hassouna, I. (2018). Behind the Number: A review of index methodologies to improve innovation measurement in Africa. Working Paper 11 Access to Knowledge for Development Center (A2K4D). American University in Cairo (February 8).

Hebb, D. (2002[1949]. *The Organization of Behavior*. Online: Psychology Press.

Hecht-Nielsen, R. (1987). Theory of the back propagation neural network. In Proceedings of the International Joint Conference on Neural Networks (pp. 1:593-608). San Diego: SOS Printing.

Herbrich, R., Keilbach, M., Graepel, T., Bollmann-Sdorra, P., & Obermayer, K. (1998). Neural networks in economics: Background, applications, and new developments. Department of Computer Science. Technical University of Berlin, Germany.

Hidalgo, C.A. (2018). The rise of knowledge economics. Scientific American (December 13). https://blogs.scientificamerican.com/observations/the-rise-of-knowledge-economics/. Accessed May 15, 2019.

Hidalgo, C.A., & Hausmann, R. (2009). The building blocks of economic complexity. *PNAS*, 106(26): 10570-10575.

Hidalgo, C.S, Klinger, B., Barabasi, A.L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science* 317 (482). Doi: 10.1126/scince.1144581.

Hinton, G. (1987). Learning translation invariant recognition in massively parallel networks. In Proceedings Conference on Parallel Architectures and Languages Europe, pp. 1-13. Springer Verlag.

Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(02): 107-116.

Honaker, J., & King, G. (2010). What to do about missing values in time-series cross-section data?. *American Journal of Political Science*, 54 (2): 561-581

Hopfield, J.J. (1982). Neural networks and physical systems with emergent collective computational abilities, Proceedings of the National Academy of Sciences of the USA, 79 (8): 2554–2558 (April).

Hopfield, J.J., & Tank, D. (1986). Computing with neural circuits. *Science* 233: 625- 633.

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks* 2(5): 359-366.

Hush, D.R., & Horne, B.G. (1993). Progress in supervised neural networks. *IEEE signal processing magazine*, 10(1), 8-39.

Jaffe, A.B., Trajtenberg, M., & Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, 108(3): 577-598.

Johnson, R., &Zhang, T. (2013). Accelerating stochastic gradient descent using predictive variance reduction. In *Advances in Neural Information Processing Systems* (pp. 315-323)

Karsoliya, S. (2012). Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture. *International Journal of Engineering Trends and Technology*, 3(6): 714-717.

Kennedy, P. (1992). *A Guide to Econometrics*. Third Edition. Cambridge (MA): The MIT Press. Chapter 14.

Khayyat, N.T., & Lee, J-D. (2012). A new measure of technological capabilities for developing countries. TEMEP Discussion Paper No. 2012:91. Technology Management, Economics, and Policy Program, College of Engineering, Seoul National University.

Kinne, J., & Lenz, D. (2019). Predicting innovative firms using web mining and deep learning, ZEW Discussion Papers, No. 19-001, ZEW - Leibniz-Zentrum für Europäische Wirtschaftsforschung, Mannheim.

Kleinberg, J., Lakkaraju, H.., Leskovec, J., &Mullainathan, S. (2018). Human decisions and machine predictions. *The Quarterly Journal of Economics*, 133(1): 237–293.

Kuan, C.M., & Liu, T. (1995). Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics* 10: 347-364.

Kuhlman, C., Ramamurthy, K.N., Sattigeri, P., Lozano, A.C., Cao, L., Reddy, C., Mojsilovic, A., & Varshney, K.R. (2017). How to Foster Innovation: A Data-Driven Approach to Measuring Economic Competitiveness. *IBM J. Res. Dev*. 61(6): 11:1-11:12.

Kuhn, M. (2015). A short introduction to the caret package. R Found Stat Comput: 1-10.

Landes, D. (1998) *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor*. New York: W.W. Norton

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature* 521(7553): 436.

Lim, H.K., Narrisetty, N.N., & Cheon, S. (2017). Robust multivariate mixture regression models with incomplete data. *Journal of Statistical Computation and Simulation*, 87(2): 328-347.

Maladkar, K. (2018). 6 types of artificial neural networks currently being used in machine learning. (January 15); https://www.analyticsindiamag.com/6-types-of-artificial-neural-networks-currently-being-used-in-todays-technology/. Accessed May 15, 2019.

McCulloch, W.S., & Pitts, W. (1943). *A logical calculus of the ideas immanent in nervous activity. Bulletin of Mathematical Biophysics*, 5:115–133.

McDonough, I.K., & Millimet, D.L. (2016). Missing data, imputation, and endogeneity. *Journal of Econometrics*, 199(2): 141-155.

McNelis, P., & McAdam, P. (2005). Forecasting inflation with thick models and neural networks. *Economic Modelling*, 22(5): 848-867.

Mihaylova, I. (2018). Applications of Artificial Neural Networks in Economics and Finance. In Encyclopedia of Information Science and Technology, Fourth Edition. Chapter 5755: 6631-6641. https://www.igi-global.com/chapter/applications-of-artificial-neural-networks-in-economics-and-finance/184358.

Miller, J.I. (2009). Cointegrating regressions with messy regressors: Missingness, mixed frequency, and measurement error. Department of Economics, University of Missouri.

Moghaddam, A.H., Moghaddam, M.H., & Esfandyari, M. (2017). Stock market index prediction using artificial neural network. *Journal of Economics, Finance and Administrative Science*, 21(41): 89-93.

Moshiri, S., & Cameron, N. (2000). Neural network versus econometric models in forecasting inflation. *Journal of Forecasting*, 19: 201-217.

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, 31(2): 87–106.

Nair, H., Kumar, A., & Ahmed, O. (2014). Neural Network Modelling, Simulation and Prediction of Innovation Growth in United Arab Emirates (UAE). *Procedia Computer Science*, 3: 269-275

Nakamura, E. (2005). Inflation forecasting using a neural network. *Economics Letters*,86: 373-378.

Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography*, 87 (3): 237-265.

Nobel Prize Committee 2018 Integrating nature and knowledge into economics. The Prize in Economic Sciences2018. https://www.nobelprize.org/uploads/2018/10/popular-economicsciencesprize2018.pdf. Accessed May 15, 2019.

North, D.C. (1990). *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press

Nowak, A., Vallacher, R.R., & Burnstein, E. (1998). Computational social psychology: A neural network approach to interpersonal dynamics. In Computer Modeling of Social process. Edited by Win B.G. Liebrand, Andrzej Nowak and Rainer Hegselman. London: Sage Publications. Chapter 6

Ojanpera, S., Graham, & Zook, M. (2019). The digital knowledge economy index: Mapping content production. *Journal of Development Studies*. https://doi.org/10.1080/00220388.2018.1554208 (January 30). Accessed May 15, 2019.

Parente, S., & Prescott, E. (1994). Barriers to technology adoption and development. *Journal of Political Economy*, 102: 298-321.

Parcero, O.J., & Ryan, J.C. (2017). Becoming a Knowledge Economy: The Case of Qatar, UAE, and 17 Benchmark Countries. *Journal of Knowledge Economy*, 8, 1146–1173.

Pao, H. T., & 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.

Rao, B.C. (2010). On the methodology for quantifying innovations. *International Journal of Innovation Management*, 14 (5): 823-840.

Rizk, N., El Said, A., Weheba, N., and de Beer, J. (2018). Towards an alternative assessment of innovation in Africa. Working Paper 10. Access to Knowledge for Development Center (A2K4D). American University in Cairo. (February 8).

Romer, P.M. (1990). Endogenous technological change. *Journal of Political Economy,* 98(5): S71-S102.

Romer, P.M. (1994). The origins of endogenous growth. *Journal of Economic Perspectives,* 8(1): 3-22.

Rukšenaite, J., &bVaitkus, P. (2012). Prediction of composite indicators using combined method of extreme learning machine and locally weighted regression. *Nonlinear Analysis: Modelling and Control*, 17 (2): 238–251.

Rumelhart, D.E., G. E. Hinton, & R. J. Williams (1986). Learning representations by backpropagating errors. *Nature* 323: 533-536.

Ryan, G., & Shinnick, E. (2011). Economic incentives and the knowledge economy: In Encyclopedia of Knowledge Management, Second Edition, Chapter 23.

Simpson, P.K. (1990). *Artificial Neural Systems*. New York: Pergamon Press.

Singh, J. (2004). Collaborative networks as determinants of knowledge diffusion patterns. Harvard Business School. Harvard University. Available at https://flora.insead.edu/fichiersti_wp/inseadwp2004/2004-74.pdf

Soley-Bai, M. (2013). Dealing with missing data: key assumptions and methods for applied analysis. Technical Report No. 4. Boston University, School of Public Health.

Sutz, J. (2012). Measuring innovation in developing countries: Some suggestions to achieve more accurate and useful indicators. *International Journal for Technological Learning and Development*, 5(1/2): 40-55.

Swanson, N., &White, H. (1997). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics*, 79 (4): 540-550.

Tchamyou, S.V. (2016). The role of knowledge economy in African business. *Journal of the Knowledge Economy*, DOI: 10.1007/s13132-016-0417-

Tibshirani, R., Wainwright, M., & Hastie, T. (2015). *Statistical Learning with Sparsity: The Lasso and Generalizations*. Chapman and Hall/CRC.

Tkacz, G., & Hu, S. (1999). Forecasting GDP Growth Using Artificial Neural Networks. Working Paper 99-3/ Document de travail 99-3.

Tzafestas, S. G., Dalianis, P. J., &Anthopoulos, G. (1996). On the overtraining phenomenon of backpropagation neural networks. *Mathematics and Computers in Simulation*, 40(5-6), 507-521.

Vapnik, V. (1982). *Estimation of Dependences Based on Empirical Data*. New York: Springer-Verlag.

Vapnik, V. (1995). *The Nature of Statistical Learning Theory*. New York: Springer-Verlag.

Vapnik, V. (1998). *Statistical Learning Theory*. New York: John Wiley and Sons.

Vapnik, V. & A. Chervonenkis (1971). On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Application*, 16(2): 264-281.

Vapnik, V., & Chervonenkis, A. (1964). A note on one class o perceptrons. *Automation and Remote Control*, 25.

Vapnik, V., & Lerner, A. (1963). Pattern recognition using generalized portrait method. *Automation and Remote Control,* 24: 774-780.

Varian, H.R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, 28(2), 3-28

Verkooijen, W. (1996). A neural network approach to long-run exchange rate prediction. *Computational Economics*, 9: 51-65.

Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10): 1550-1560.

White, H. (1988). Economic Prediction Using Neural Networks: The Case of the IBM Daily Stock Returns. Proceedings of IEEE International Conference on Neural Networks, San Diego, 24-27 July 1988, 451-458. http://dx.doi.org/10.1109/ICNN.1988.23959.

World Bank Institute. Measuring knowledge in the world's economies: Knowledge assessment methodology and knowledge economy index. http://web.worldbank.org/archive/website01030/WEB/IMAGES/KAM_V4.PDF. Accessed May 15, 2019.

Zanak, S.Z., & Becerra-Fernández, I. (2005). Competitiveness of nations: A knowledge discovery examination. *European Journal of Operations Research*, 166: 185-211

Zaremba, W., Sutskever, I., & Vinyals, O. (2014). Recurrent neural network regularization. arXiv preprint arXiv:1409.2329.

Zeiler, M.D. (2012). ADADELTA: An adaptive learning rate method. arXiv preprint arXiv:1212.5701.

Zhang, S. (2012). Nearest neighbor selection for iteratively kNN imputation. *Journal of Systems and Software*, 85(11): 2541-2552.

Zhu, G., Zhaoxiang, Zhang, Xu-Yao, & Liu, Cheng-Lin (2017). Diverse neuron type selection for convolutional neural networks. Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17): 3560-3566.

Zurada, J.M. (1992). *Introduction to Artificial Neural Systems* (Vol. 8). St. Paul: West publishing company.

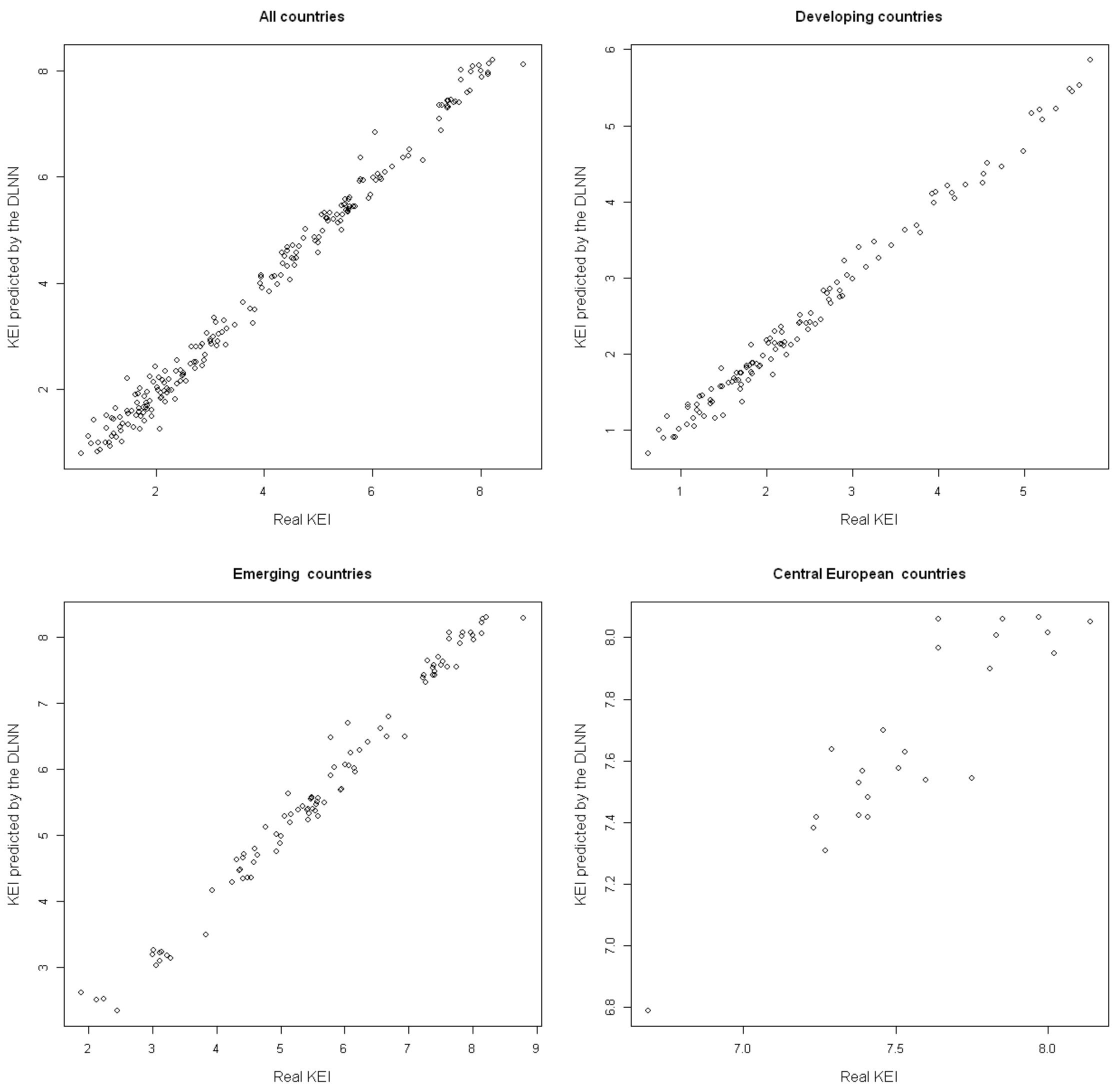
All countries
Developing countries
Emerging countries
Central European countries
KEI predicted by the DLNN
Real KEI

**Table 1**
***Variable Definitions, and Data Sources***

| ***Variable*** | ***Definition*** | ***Source*** |
|---|---|---|
| ***Regulatory quality (REGQU)*** | This indicator measures the incidence of market-unfriendly policies such as price controls or inadequate bank supervision, as well as perceptions of the burdens imposed by excessive regulation in areas such as foreign trade and business development | Worldwide Governance Indicators (WGI). The World Bank. Available at https://info.worldbank.org/governance/wgi/#home |
| ***Rule of law (RULEL)*** | This indicator includes several indicators that measure the extent to which agents have confidence in and abide by the rules of the society} | Worldwide Governance Indicators (WGI). The World Bank. Available at https://info.worldbank.org/governance/wgi/#home |
| ***Tariff and non-tariff barriers (TNTBA)*** | This is a score assigned to each country based on the analysis of its tariff and non-tariff barriers to trade, such as import bans and quotas as well as strict labeling and licensing requirements | The Heritage Foundation's Trade Freedom score. The Heritage Foundation. Available at https://www.heritage.org/index/trade-freedom |
| ***Patent applications (PATENT)*** | Patent grants by country of origin and patent office per 1000 people | World Development Indicators. Data Bank. The World Bank. Available at https://data.worldbank.org/ |
| ***Scientific and technical journal articles (STJOU)*** | Scientific and engineering articles published by country per 1000 people | World Development Indicators. Data Bank. The World Bank. Available at https://data.worldbank.org/ |
| ***Internet users (INTERN)*** | Internet users are individuals who have used the Internet (from any location) in the last 3 months per 1000 people | World Development Indicators. Data Bank. The World Bank. Available at https://data.worldbank.org/ |
| ***Fixed telephone subscriptions (TELEP)*** | The number of subscriptions per 1000 people. It includes Integrated services digital network channels and fixed wireless subscribers | World Development Indicators. Data Bank. The World Bank. Available at https://data.worldbank.org/ |
| ***Fixed broadband internet subscribers (FIXBI)*** | Fixed broadband internet subscribers per 1000 people | World Development Indicators. Data Bank. The World Bank. Available at https://data.worldbank.org/ |
| ***Primary enrolment (% gross) (PRIMARY)*** | Gross enrollment ratio is the ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the primary level of education | World Development Indicators. Data Bank. The World Bank. Available at https://data.worldbank.org/ |
| ***Secondary enrolment (% gross) (SECONDARY)*** | The ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the secondary level of education | World Development Indicators. Data Bank. The World Bank. Available at https://data.worldbank.org/ |
| ***Tertiary enrolment (% gross) (TERTIARY)*** | The ratio of total enrolment, regardless of age, to the population of the age group that officially corresponds to the tertiary level of education | World Development Indicators. Data Bank. The World Bank. Available at https://data.worldbank.org/ |
| ***Total Population (POP)*** | Total population. Mid- year estimates | World Development Indicators. Data Bank. The World Bank. Available at https://data.worldbank.org/ |

Table 2. Basic descriptive statistics

| ***Variable*** | ***Mean*** |
|---|---|
| ***Patent applications*** | 2.24 |
| ***Fixed broadband internet subscribers*** | 82.99 |
| ***Tertiary enrolment (%)*** | 33.49 |
| ***Scientific journal articles*** | 2.03 |
| ***Secondary enrolment (%)*** | 77.17 |
| ***Regulatory quality*** | -0.0167 |
| ***Rule of Law*** | -0.0720 |
| ***Primary enrolment (%)*** | 101.02 |
| ***Tariff and non-tariff barriers*** | 69.02 |
| ***Internet users*** | 24.05 |
| ***Telephone lines*** | 181.04 |
| ***Total population*** | 37940000 |

Table 3. Distribution of missing values

| *Variable* | *Number of Missing Cases* | *% of Missing Cases* |
|---|---|---|
| ***Patent applications*** | 759 | 76.7 |
| ***Fixed broadband internet subscribers*** | 450 | 45.5 |
| ***Tertiary enrolment*** | 420 | 42.5 |
| ***Scientific journal articles*** | 392 | 39.6 |
| ***Secondary enrolment*** | 370 | 37.4 |
| ***Regulatory quality*** | 172 | 17.4 |
| ***Rule of Law*** | 172 | 17.4 |
| ***Primary enrolment*** | 172 | 17.4 |
| ***Tariff and non-tariff barriers*** | 131 | 13.2 |
| ***Internet users*** | 106 | 10.7 |
| ***Telephone lines*** | 15 | 1.52 |
| ***Total population*** | 6 | 0.607 |