USING ARTIFICIAL NEURAL NETWORKS FOR INCOME CONVERGENCE

Kayhan Koleyni, Islamic Azad University of Khorasgan (Isfahan)

ABSTRACT

Economic convergence is an important topic in modern Macroeconomics. Economic convergence refers to the tendency of per capita income of countries (regions) to approach their steady-state value. Two types of convergence are identified in the literature: Conditional and Absolute Convergence. This paper studies income convergence between 177 world countries during the period of 1980-2006 by using the neoclassical growth model of Barro-Sala-i-Martin for both kinds of convergence. Non-linearity of the underlying relationships, the restrictiveness of assumptions of functional forms and econometric problems in the estimation and application of theoretical models, advocate for the use of Artificial Neural Networks (ANN) algorithms. We show that by changing the quantitative tools of analysis and using ANN results become more precise. Results show that absolute convergence does not exist and conditional convergence is insignificant

JEL: C45, E37, O47

KEYWORDS: Economic convergence, non-linearity, econometrics, artificial Neural Networks algorithms

INTRODUCTION

The movement of the world towards integration, polarization and the formation of united countries is one that requires attention. Economically, debates about integrations such as European, Islamic, G7 or ASEAN are studied in different fields such as international economics, growth economics, etc. However, one of the concepts, which is essential to all these fields of study is income (output) convergence. Income convergence is one of the topics of new Macroeconomics. It refers to the tendency of per capita income of countries (regions) to converge their steady-state value.

Convergence hypothesis tries to answer two main questions. First, do poor countries (regions) grow faster than rich ones? This attempts to consider the effect of initial conditions on per capita income differences across countries, and the speed of convergence, which is introduced by β-convergence in growth literature. Second, does the dispersion of per capita income of countries (regions) decrease over time. This type of convergence called σ-convergence, focuses less on initial conditions and instead emphasizes income distribution by measuring standard deviation.

When studying β-convergence, two kinds of convergence should be studied: conditional and absolute convergence. If the differences in per capita income are temporary and solely because of initial conditions absolute convergence is occurring. If the differences are permanent because of cross-country structural heterogeneity, conditional convergence is occurring (Durlauf et al., 2005). Expansion of the literature on economic growth, its modeling and the development of additional quantitative tools of analysis and different type of statistical information used for quantitative analysis (cross-section, time series or panel data) have promoted a large body of empirical studies about convergence hypothesis. However, criticism of both kinds of β and σ convergence still exists. According to Durlauf et al. (2005), one of the criticisms about β-convergence is the effects of linear approximation. There is a body of research that explores the effects of the approximations that are employed to produce the models used to evaluate β- convergence. Durlauf and Johnson (1995), Binard and Pesaran (1999) and Liu and Stengos

(1999) presented evidence against the adequacy of linear approximation. On the other hand, Romer (2001) and Dowrick (2004) claimed that the approximation was quite reliable. Accordingly, some of these studies show the accuracy of linear approximation while others do not. In addition, Durlauf et al. (2005) declared that nonlinearity had a deeper affect than simple approximation error and could affect the steady state of per capita income and its identification problem, which is another criticism of β-convergence.

We try to solve criticisms about the non-linearity of the underlying relationships and the ambiguity of functional form by using a non-parametric approach: Artificial Neural Networks (ANN). We show that by changing, the quantitative tools of analysis from traditional econometric tools to the new ANN approach results become more precise, the non-linearity problem is solved and the appropriate functional form of movements of the per capita income of countries toward their steady-state value is identified. We used Multilayered Feed-Forward Networks for this purpose and compared it with the Ordinary Least Squares (OLS) estimation method based on the cross-country regression equations of Barro and Sala-i-Martin model (1992, 1995 and 2004) for both kinds of absolute and conditional β-convergence during the period of 1980-2006. This paper is organized as follows: in section 2, we discuss some main studies on convergence hypothesis; in section 3, we discuss the methodology and ANN modeling; in section 4, we describe our findings and compare them with the OLS method; finally, in section 5, we draw our conclusions.

LITERATURE REVIEW

As mentioned earlier, a large number of studies have been performed using different types of data, countries, sample periods and choice of control variables, but we describe here only small portion of this large body of empirics. Many of these empirical studies are based on neoclassical models of growth such as Solow (1956), Swan (1956), Koopmans (1965), Cass (1965) and even the previous work of Ramsey (1928). In addition, we see some of studies with endogenous models of growth as described by Jones and Manuelli (1990) and Kelly (1992). However, most endogenous models of growth are not compatible with the convergence hypothesis, such as the model described by Romer (1986) and Lucas (1988) because of the convexity in the production function.

Convergence hypothesis originates from Abramovitz (1986) and Baumol (1986). Baumol used data from 1870-1974 for 16 OECD countries and estimated the regression shown in equation (1). He concluded the perfect convergence occurs (b is approximately near -1). However, in a later study, Delong (1988) showed that Baumol’s conclusions were not correct because of problems with sample selection and measurement error.

$$\ln\left({}^{y}/_{N}\right)_{i,1979} - \ln\left({}^{y}/_{N}\right)_{i,1870} = a + b \ln\left[\frac{y}{N_{i,1870}}\right] + \varepsilon_i \tag{1}$$

Where, y is the per capita income of countries, N denotes the number of countries, and ε_i is the error term. Barro and Sala-i-Martin (1992) defined β and σ-convergence for US states according to the Solow model. They developed the cross-country regression shown below (equation 2):

$$\log\left(\frac{\mathrm{y_{it}}}{\mathrm{y_{i,t-1}}}\right) = a_{it} - \left(1 - e^{-\beta}\right).\log\left(y_{i,t-1}\right) + u_{it} \tag{2}$$

Where, the subscript t denotes the year, the subscript i denotes the country or region, y is the per capita income, and u is the error term. If we assume that coefficient a_{it} is the same for all economies, then $a_{it} = a_t$. This specification means that the steady state value and the rate of exogenous technological progress are the same for all economies. This assumption is more reasonable for regional data sets than across world countries. It is plausible that different regions within a country are more similar than different countries across the world with respect to technology and preferences. As most of the research shows, global absolute convergence does not exist. If the intercept a_{it} is the same for all places, and $\beta > 0$, then equation (2) implies that poor economies tend to grow faster than rich ones. This type of convergence is called "absolute" or "unconditional" convergence.

If one uses the term $\left(1-e^{-\beta}\right).\log(\hat{y}_i^*)$ as an explanatory variable, it means that the growth rate of economy i depends on its initial level of income and also depends on the steady state value of income, $\hat{y}_i^*$. This is why we use the concept of "conditional" rather than absolute convergence. The growth rate of an economy depends negatively on its initial level of income, after conditioning the steady state. It is as follows:

$$\log\left(\frac{y_{it}}{y_{i,t-1}}\right) = a_{it} - \left(1-e^{-\beta}\right).\log\left(y_{i,t-1}\right) + \left(1-e^{-\beta}\right).\log(\hat{y}_i^*) + u_{it} \qquad (3)$$

This study is the basis for all other studies using the cross-section approach and β convergence. The authors also analyzed the pattern of β convergence for Japanese prefecture, across European regions, etc. Despite their attempts to improve errors related to their definition, criticisms remain about measurement error, endogeneity and effects of linear approximation.

Mankiw, Romer and Weil (1992) used the augmented Solow model, which includes accumulation of human as well as physical capital for several large groups of countries, and found similar results to Barro and Sala-i-Martin (1992). There is a large body of research based on the findings of Barro and Sala-i-Martin (1992) and the cross-section approach. We also, see other efforts in different approaches such as time series and panel data analysis. Bernard and Durlauf (1995, 1996), Durlauf (1998), a series of Quah's papers (1992, 1993a, 1993b, 1996a, 1996b, 1996c, 1997, 2001) and Nahar and Inder (2002) tried to use time series approach in their studies. This approach is largely statistical in nature. A disadvantage of this approach is that it is not in accordance with any particular growth theory. Lee, Pesaran and Smith (1997), Islam (1995), Caselli, Esquivel and Lefrot (1996), Benhabib and Siegel (1997), Nerlove (1996), Conva and Marcet (1993) and Evans (1998) used the panel data approach. Panel data analysis adapts the use of convergence equations as done in cross-section analysis. Although, it can help to increase the flexibility of model, structural error can be observed with this approach.

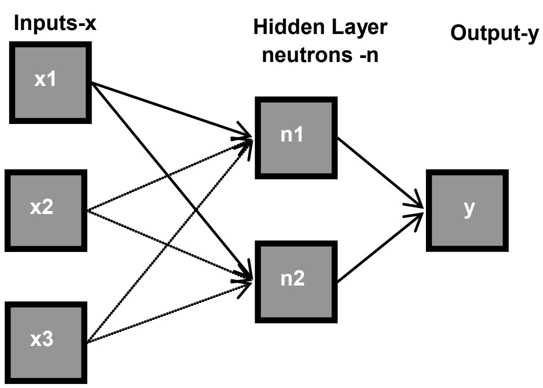
Papadas and Estratoglou (2004) tried to analyze β convergence through the cross-section approach in accordance with the Barro and Sala-i-Martin model. They used the concept for 52 prefectures of the Greek economy over two periods, according to availability of investment data. They estimated both conditional and absolute convergence for the periods of 1981-1991 and 1971-1991. Additional variables for analyzing conditional convergence were the percentage share of the total labor force employed in the primary sector, the percentage share of the total population with secondary education, investment and unemployment rate. The authors introduced the artificial neural networks (ANN) algorithm as a useful tool for studying the non-linearity relationship of β convergence. They utilized a Back-Propagation Network (BPN) with 10 neurons and 1 bias node and showed that it can perform very well and more accurately. According to Papadas and Estratoglou (2004) there had been no other study of ANN application to the empirics of convergence, their study being the first. Although, their study introduced the ANN algorithm, lack of analysis of substitution neural networks remains the most important criticism of their research. Efficiency of neural networks is every much related to the architecture and design of these networks. Different networks with different architectures should be designed and among them, the best network with the lowest error should be chosen. In addition, the length of the studied periods for neural networks is the other topic for discussion. Usually, neural networks with longer periods are more accurate.

METHODOLOGY

Artificial neural networks are members of a family of statistical techniques, which try to simulate and model the human brain. They have recently received a great deal of attention in many fields of study. A neural network relates a set of input variables (input layers) to a set of one or more output variables (output layers). The component of each layer is called a neuron or node. The difference between a neural network and other approximation methods is that neural networks make use of one or more hidden layers, in which the input variables are transformed by a special function in parallel processing. Each neuron has one ascendant activation function, which can be linear or nonlinear according to their application. This activation function determines the threshold of the neuron. The neuron receives a weighted sum of inputs from a connected unit, and reply according to this threshold and a weighted sum of inputs. The threshold behavior of *logsigmoid* and *tansig* or *tanh* activation functions, which characterizes many types of economic responses to changes in fundamental variables, explains their significant adoption in the economy.

This section identifies two different neural networks: feed-forward networks with Back-Propagation (BPN) learning algorithm, mostly used by economists for prediction and Elman Recurrent networks. Figure 1 illustrates the architecture of feed-forward networks.

Figure 1: Architecture of Feed-Forward Networks



This figure shows the architecture of feed-forward back propagation networks. Inputs X makes the first layer of the networks. After this layer, hidden layer with n neurons processes the inputs in parallel. Final layer of a network is output layer.

The source nodes in the input layer of the feed-forward network supply respective elements of the activation pattern (input vector), which constitute the inputs applied to the neurons (computation nodes) in the second layer (i.e., the first hidden layer). The outputs of the second layer are used as inputs of the third layer, and so on for the rest of the network. These networks can be connected fully or partially. These networks have the ability to learn from the environment and dataset, and improve their performance through learning; the improvement in performance takes place over time in accordance with some prescribed measure. A neural network learns about its environment through an iterative process of adjustments applied to network's weights and thresholds. Ideally, the network becomes more knowledgeable about its environment after each iterate of learning process. Two kinds of learning processes exist: supervised learning and unsupervised learning. The back-propagation algorithm has emerged as the most popular algorithm for the supervised learning of multilayer feed-forward networks.

The following system represents the multilayer feed-forward network:

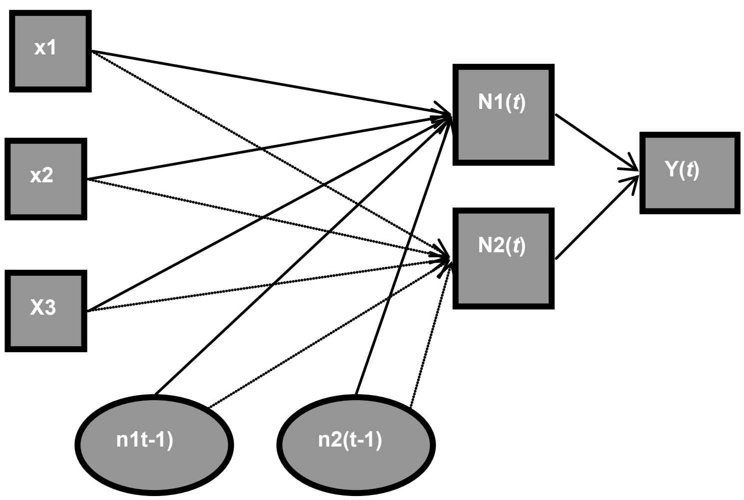
$$n_{k,t} = \omega_{k,0} + \sum_{i=1}^{i^*} \omega_{k,i} x_{i,t} \tag{4}$$

$$N_{k,t} = L(n_{k,t}) = \frac{1}{1+e^{-n_{i,t}}} \tag{5}$$

$$y_t = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k N_{k,t} \tag{6}$$

Where, $L(n_{k,t})$ represents the *logsigmoid* activation function with the form $\frac{1}{1+e^{-n_{i,t}}}$. In addition, the alternative activation function, which is known as *tansig* or *tanh* with the form, $\frac{e^{n_{k,t}}-e^{-n_{k,t}}}{e^{n_{k,t}}+e^{-n_{k,t}}}$ could be used. In this system, there are i^* input variables $\{x\}$, and k^* neurons. A linear combination of these input variables observed at time *t*, $\{x_{i,t}\}, i = 1, \dots, i^*$, with the coefficient vector or set of input weights $\omega_{k,i}, i = 1, \dots, i^*$, as well as the constant term $\omega_{k,0}$, form the variable $n_{k,t}$. This variable is transformed by the activation function, and becomes a neuron $N_{k,t}$ at time or observation *t*. The set of k^* neurons at time or observation index *t* are combined in a linear way with the coefficient vector $\{\gamma_k\}, = 1, \dots, k^*$, and taken with constant term γ_0 , to form the forecast $\hat{y}_t$ at time *t*. A recurrent network distinguishes itself from a feed-forward neural network in that it has at least one feedback loop. Figure 2 illustrates the architecture of Elman recurrent network.

Figure 2: Architecture of Elman Recurrent Network



This figure shows the architecture of Elman Recurrent networks. Inputs X makes the first layer of the networks. After this layer, hidden layer with n neurons processes the inputs in parallel. Final layer of a network is output layer. This network has a feedback from the hidden layer that works like a memory for network and make the network dynamic.

This network allows the neurons to depend not only on the input variables *x*, but also on their own lagged values. Thus, the Elman network builds "memory" in the evolution of neurons. The following system represents the recurrent Elman network illustrated in figure 2:

$$n_{k,t} = \omega_{k,0} + \sum_{i=1}^{i^*} \omega_{k,i} x_{i,t} + \sum_{k=1}^{k^*} \emptyset_k n_{k,t-1} \tag{7}$$

$$N_{k,t} = \frac{1}{1+e^{-n_{i,t}}} \tag{8}$$

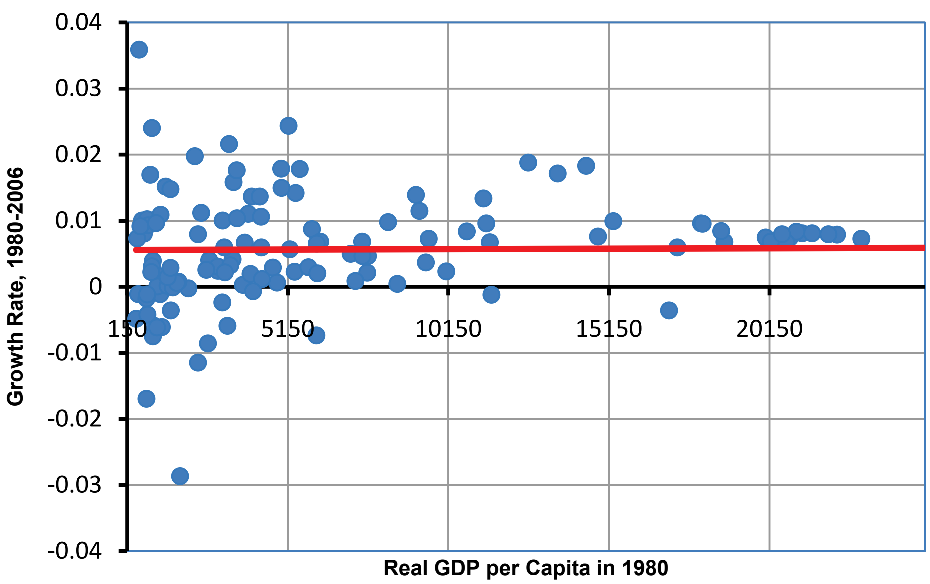
$$y_t = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k N_{k,t} \tag{9}$$

Note that the recurrent Elman network is one in which the lagged hidden layer neurons feedback into the current hidden layer of neurons. Unlike the feed-forward, which is a static network, Elman is dynamic because it uses its feedback and the network's state changes up to the time that network consider the steady state.

ANALYSIS OF RESULTS

Figure 3 illustrates the growth rates of real GDP per capita from 1980 to 2006 against the levels of real GDP per capita in 1980. The red line is the straight line that provides a best fit for the relationship between the growth rate of real GDP per capita (the variable on the vertical axis) and the level of real GDP per capita (on the horizontal axis). If the convergence prediction from the Solow model is correct, we should find low levels of real GDP per capita matched with high growth rates, and high levels of real GDP per capita matched with low growth rates. Instead, it is difficult to discern any pattern in the data; if anything, there is a very slight tendency for the growth rate to rise with the level of real GDP per capita. In addition, we can see that most of countries are concentrated on the left side of the plot, showing that real GDP per capita for these countries is less than $10,000. Based on these facts, we test convergence hypothesis with equations (2) and (3) from Barro and Sala-i-Martin model. Next, the ANN algorithm is used to make β convergence estimation more accurate, which solves the nonlinearity problem of the definition. We used the GDP per capita index from the World Bank 2008 database for the analysis.

Figure 3: Growth Rate versus Level of Real GDP per Capita for a Board Group of Countries



This figure shows a board group of countries with different real GDP per person in 1980 versus growth rate 1980-2006 on a proportionate scale. The red line is the trend line between countries data. If the slope of this trend line be negative means that convergence exists but as we can see the slop is almost positive.

Table 1 shows the estimation results of equation (2) for 177 countries during the period of 1980-2006. It includes the estimated value of the coefficient of the independent variable with its *t*-value in parentheses, the corresponding derived value of β and the value of R^2. The estimated negative value of β, derived from the positive coefficient of $log\ (y_{i,t_0})$, which is statistically significant, demonstrates a lack of absolute β convergence. The relatively low value of R^2 is not unusual in such cross-sectional equation estimates. Such values in general can reflect the significance of omitted factors. In addition, structural heterogeneity and differences in the initial conditions of the 177 countries may determine different steady states for these countries.

Table 1: Results of Absolute Convergence Regression

β	Estimated coefficient of $\log(y_{i,t_0})$	R^2
-0.0013	0.0013 (2.55)	0.0015

This table presents the results of linear regression model for absolute convergence, the negative sign of β support absence of absolute convergence.

Equation (10) summarizes these results, Where gy is the annual growth rate of country *i* at time *t*:

$$\widehat{gy}_{it} = -0.0177 - (1 - e^{0.0013}).\log(y_{i,t-1}) \tag{10}$$

Accordingly, after we condition on steady state as described in equation (3) by adding different independent variables, we can estimate the conditional convergence. Table 2 presents the conditional convergence results. Variables added linearly to the original model include openness in constant price (OPEN), total labor force (LAB), investment share of real GDP (INV), and percentage of government expenditure from GDP (GOV).

Table 2: Results of Conditional Convergence Regression

β	$\log(y_{i,t_0})$	OPEN	LAB	INV	GOV	R^2
-0.00067	0.00067 (1.23)	0.000046 (3.47)	----	----	----	0.0047
-0.00065	0.00065 (1.19)	0.000051 (3.76)	0.0000013 (1.23)	----	----	0.0055
-0.0011	0.0011 (1.76)	0.000048 (3.59)	----	-0.00014 (-1.44)	----	0.0052
-0.00088	0.00088 (1.54)	0.000045 (3.40)	----	----	-0.00020 (-2.05)	0.0053
-0.0015	0.0015 (2.27)	0.000055 (3.93)	0. 0000017 (1.55)	-0.0002 (-1.97)	-0.00021 (-1.92)	0.0069

This table presents the results of linear regression model for conditional convergence, the negative sign of β support absence of absolute convergence. T-statistics show that only openness is significant and the results do not support significance of conditional convergence.

As the results show, additional variables improve the explanatory power of the model very slightly, and in all presented models of the table 2 coefficient of $\log(y_{i,t_0})$ is statistically insignificant. With the exception of the coefficient of openness, which is statistically significant and has a very small value, all other additional variables are statistically insignificant and add nothing to the model. It seems that adding additional variables to the model if they are statistically significant, only improves the model's explanatory power and the insignificancy of $\log(y_{i,t_0})$ still remains. As previous studies and the facts of figure 3 show, convergence on both levels of conditional and unconditional for the period of 1980-2006 does not exist, and the hypothesis of income convergence is rejected across 177 world countries.

Next, we use the nonparametric ANN approach to improve the accuracy of the estimation. A neural network uses three samples of data. The training sample is presented to the network during the learning and training process and the network is adjusted according to its error. The validation sample is used to measuring network generalization and to halt training when generalization stops improving. The testing sample has no effect on training, and so provides an independent measure of network performance during and after training. The most important challenge of neural network performance is related to its architecture. The standard method of designing a suitable network is trial and error. We examined more than 50 different neural networks to find the best network's architecture empirically and used training, validation and testing subsamples with 70%, 15%, 15% - 65%, 15%, 20% and 60%, 20%, 20% (which are general orders for ANN algorithm) and found that the 70%, 15%, 15% subsample is best. Therefore, we chose 1980-98 as the training sample, 1998-2002 as the validation sample and 2002-2006 as the testing sample. Considering equations (4) to (9) we used the BPN feed-forward network model as follows:

$$\log\left(\frac{y_{it}}{y_{i,t-1}}\right) = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k \left(\frac{1}{1+e^{-(\omega_{k,0}+\sum_{i=1}^{i^*} \omega_{k,i} * \log(y_{i,t-1}))}}\right) \quad (11)$$

And the ELMAN recurrent network model:

$$\log\left(\frac{y_{it}}{y_{i,t-1}}\right) = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k \left(\frac{1}{1+e^{-n_{i,t}}}\right) \quad (12)$$

$$n_{k,t} = \omega_{k,0} + \sum_{i=1}^{i^*} \omega_{k,i} * \log(y_{i,t-1}) + \sum_{k=1}^{k^*} \emptyset_k n_{k,t-1} \quad (13)$$

Two-layer networks with *tansigmoid* activation function in the hidden layer and *purelin* activation function in the output layer for both kinds of networks were used. The networks were trained with Levenberg-Marquardt back propagation algorithm (*trainlm*) for 1000 epochs and their performance was evaluated using mean squared error (MSE). Table 3 summarizes some of these networks.

Table 3: Results of Different Artificial Neural Networks

Number of Network	Network Type	Number of Hidden Neurons	MSE-TRAIN	MSE-TEST
1	BPN Feed-Forward	2	0.00071	0.00067
2	BPN Feed-Forward	3	0.00046	0.00077
3	BPN Feed-Forward	4	0.00067	0.00058
4	BPN Feed-Forward	5	0.001	0.0011
5	BPN Feed-Forward	6	0.0012	0.0014
6	BPN Feed-Forward	7	0.0013	0.0020
7	BPN Feed-Forward	10	0.0013	0.0032
8	BPN Feed-Forward	15	0.00092	0.0018
9	BPN Feed-Forward	20	0,00090	0.0020
10	BPN Feed-Forward	25	0.001	0.0025
11	BPN Feed-Forward	30	0.00055	0.0022
12	ELMAN Recurrent Network	2	0.00061	0.0011
13	ELMAN Recurrent Network	3	0.00067	0.0015
14	ELMAN Recurrent Network	4	0.00078	0.00096
15	ELMAN Recurrent Network	5	0.00062	0.00092
16	ELMAN Recurrent Network	6	0.00076	0.0014
17	ELMAN Recurrent Network	7	0.00085	0.0034
18	ELMAN Recurrent Network	10	0.0014	0.0039
19	ELMAN Recurrent Network	15	0.00079	0.0022
20	ELMAN Recurrent Network	20	0.00057	0.0025

This table illustrates different Elman and BPN feed-forward networks with different performance and different errors.MSE measurement for testing sample shows that the best capable network is the third one with 4 hidden neurons.

As the table shows, different kinds of networks with different numbers of hidden neurons produce different performance. The best network with the highest performance is chosen according to the MSE of testing sample. The ability of the network to estimate accurately the annual rates of growth, our dependent variable, based on unused values of the independent variable in training, indicates that it has sufficiently captured the underlying relationship and how well the networks perform with new data from the testing sample, which is fed to networks after training. Despite the dynamics of the Elman recurrent networks, the third neural network model BPN, feed-forward with four hidden neurons is the model that minimizes the MSE in the testing sample and produces estimations that are more accurate. Table 4 presents the performance of this network.

Table 4: Comparison of ANN and Linear Model Performance

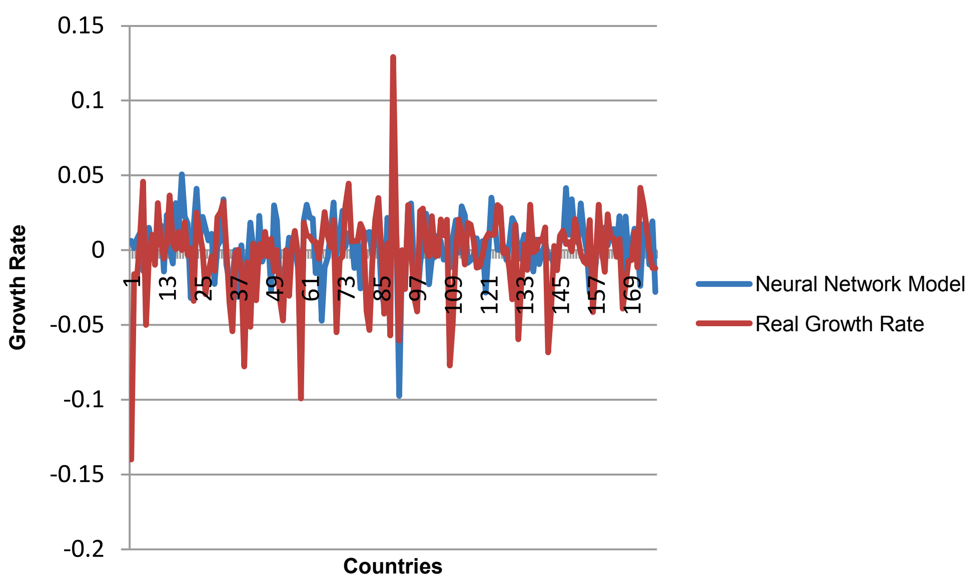
Model	MSE	MAE	MSEREG	SSE
BPN Feed-Forward	0.00072	0.0192	0.00065	3.076
Linear Regression	0.0017	0.0357	0.0015	7.202

This table shows the performance of artificial neural network and linear regression model with different measurement of their errors. It supports that ANN is more accurate.

As the results show, the mean squared error (MSE), mean absolute error (MAE), mean squared error with regularization (MSEREG) and sum squared error (SSE) of the BPN feed-forward neural network are lower than those of the linear regression model. Hence, the neural network is a more capable and more accurate model than the linear regression model.

Sometimes, the values of R^2 for ANNs are presented as well, based on the definition of the measure of fitness of data in regression analysis, but this is implausible because the definition of R^2 breaks down as we move away from traditional regression. For a better understanding of the ANN performance, figure 4 shows an in-sample evaluation. It illustrates a randomly selected year 1991, which was used during training in comparison with the real data. As the results show, the network performs well in the growth rate estimations.

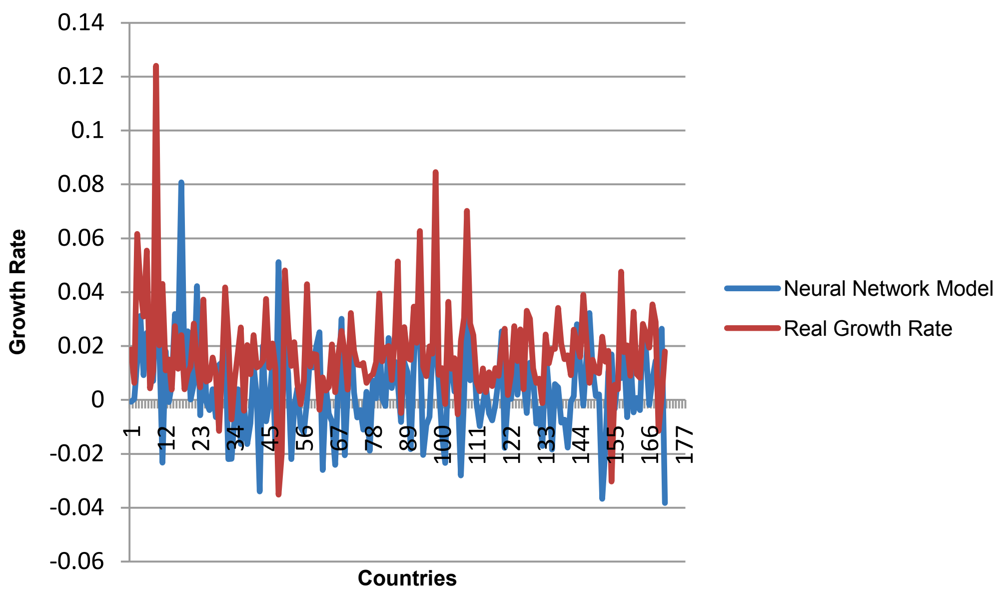
Figure 4: In-Sample Performance of ANN for 1991



This figure shows In-Sample performance of artificial neural networks for 1991, randomly selected, for 177 world countries.

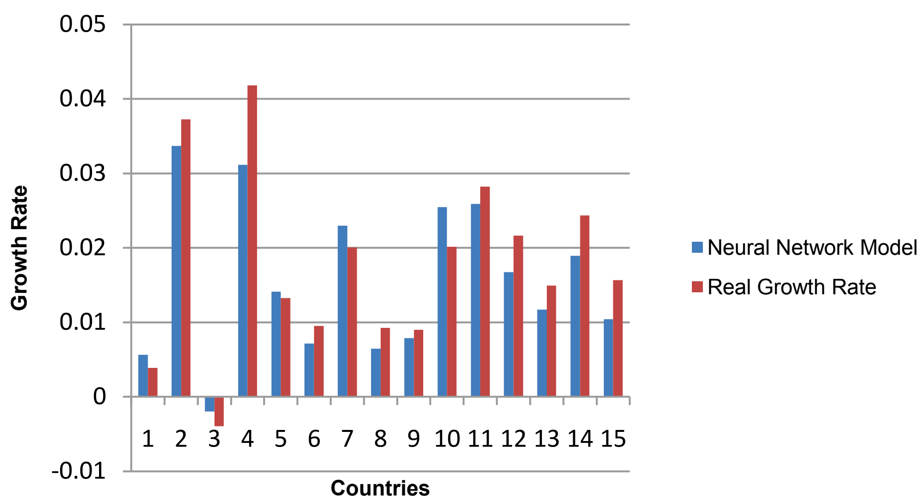
Figure 5 plots the out-of-sample performance of ANN. It shows the rates of growth of 177 countries in comparison with the estimated rates of growth of the selected neural network model for the year 2006, the last period in the testing sample. It is natural that this estimation is less accurate than in-sample estimations, but it is still appropriate. For better recognition, figure 6 illustrates the results of 15 randomly selected rates of growth of countries in histogram pattern in comparison with estimated neural network model rates.

Figure 5: Out-of-Sample Performance of ANN for 2006



This figure illustrates Out-of-Sample performance of ANN. It compares the neural network estimations with real data of growth rate of world countries in 2006.

Figure 6: Out-of-Sample Performance of ANN for 15 Countries in 2006



This histogram compare neural network results with real data of growth rate of 15 randomly selected countries in 2006. Because the histogram of 170 countries is not capable enough to show the results, we reduced it to 15 countries.

Each of these findings demonstrate the capability of neural networks for studying income convergence and capturing the movement of the growth rates of different countries. Artificial neural networks can also be used for analyzing unconditional income convergence; however, because the results of the regression model show that additional variables are not helpful and the model remains insignificant, we surrender this part.

CONCLUSIONS AND SUGGESTIONS

As our results show, absolute convergence does not exist for 177 countries across the world during the studied period of 1980-2006. This means that our analysis does not support the tendency of poor economies to grow faster than rich ones across the world. This finding is consistent with the findings of previous studies. In addition, after conditioning different variables such as openness, total labor force, investment and government expenditure, we conclude that conditional β convergence is statistically insignificant in all of our estimated models.

Nonlinearity of the underlying relationships, the restrictiveness of assumptions of functional forms and econometric problems in the estimation and application of theoretical models advocate for the use of Artificial Neural Networks (ANN) algorithms. We show that by changing the quantitative tools of analysis and using ANN the results become more precise in comparison to the OLS method. For this purpose, we examined more than 50 neural networks with different architectures in two types of feed-forward back propagations, (BPN) and the Elman recurrent. We found that despite the dynamic of the Elman recurrent networks, the BPN feed-forward neural network model with four hidden neurons produces the most accurate estimations.

The most important point of this study is that, although we compared the accuracy of neural network with regression models and concluded that the neural network is more capable, we used artificial neural networks as the complement algorithm of the OLS method, not as the alternative or substitute approach. Accordingly, when we discovered that conditional convergence is statistically insignificant, we did not continue our analysis of the ANN complement approach.

We recommend, as Barro and Sala-i-Martin (2004) mention, future studies that try to survey income convergence, use both the regression method and the ANN algorithm for similar economies, such as regions of a country or countries of a trading block. In addition, one can use artificial neural networks with other approaches, such as time series or panel data.

REFRENCES

Barro, R. & Sala-i-Martin, X (1990). "Economic Growth and Convergence across the United States", Working Paper no.3419, NBER, Cambridge. Massachusetts.

Barro, R. & Sala-i-Martin, X (1992). "Convergence", *Journal of Political Economy,* 100, 223-151.

Barro, R. & Sala-i-Martin, X (2004). *Economic Growth*. Cambridge Ma: MIT Press.

Durlauf, S. N. & Quah, D.T (1998). “The New Empirics of Economic Growth”, the University of Wisconsin, Madison and LSE.

Durlauf, S. N. & Johnson, P.A & Temple, J.R (2005). “Growth Econometric” *Handbook of Economic Growth,* Vol 1(a). p. 561-660.

Fiaschi, D. & Lavezzi, A.M(2007). “Productivity Polarization and Sectoral Dynamic in European Regions”, *Journal of Macroeconomics,* 29, 612–637.

Islam, N. (2003). "What Have We Learned from the Convergence Debate?", *Journal of Economic Surveys* ,17, 309–362.

Islam, N. & Sakamoto, H. (2006). "Convergence Across Chinese Provinces: An Analysis Using Markov Transition Matrix" *China Economic Review.*

Lee, K.& Pesaran, M. H. & Smith, R. (1997). "Growth and Convergence in a Multi-Country Empirical Stochastic Solow Model", *Journal of Applied Econometrics*, 12,357-392.

McNelis, P.D (2005). *Neural Networks in Finance: Gaining Predictive Edge in the Market.* Elsevier Academic Press.

Nahar, S. & Inder, B. (2002), "Testing Convergence in Economic Growth for OECD Countries", *Applied Economics*, 34, 2011-2022.

Pesaran, M.H. (2004a). "A Pair-Wise Approach to Testing for Output and Growth Convergence". Mimeo, University of Cambridge.

Pesaran, M.H. (2004b). "General Diagnostic Tests for Cross-Section Dependence in Panels".Mimeo, University of Cambridge.

Papadas .C.T & Efstratoglou.S (2004). "Estimation of Regional Economic Convergence Equations Using Artificial Neural Networks with Cross Section Data". Greece. Agricultural University of Athens.

Quah, D. T. (1997). "Empirics for Growth and Distribution: Polarization, Stratification, and Convergence Clubs", *Journal of Economic Growth*, 2(1), 27-59.

Solow, R. (1956). "A Contribution to the Theory of Economic Growth", *Quarterly Journal of Economics* LXX, 65-94.

Summers, R. & Heston, A. (2006). "The PennWorld Table (Mark 6.2): An expanded set of international comparisons, 1950–2004". *Quarterly Journal of Economics* 106 (2), 327–368.

W.D.I (2008). "World Development Indicator" World Bank Press.

Warner.B, Misra.M (1996). "Understanding Neural Networks as Statistical Tools" *The American Statistician*, Vol. 50, No. 4, pp. 284-293.

Yigit. T.M & Kutan, A (2003). "European Integration, Productivity Growth and Real Convergence", U.S. Southern Illinois University**.**

BIOGRAPHY

Kayhan Koleyni, M.A of economic of the Islamic Azad University of Khorasgan (Isfahan)-Iran. His field of interest includes Growth Models. He has a number of papers published in various journals. E-mail: Kayhan_Koleyni@Yahoo.com , Fax: +98-311-6639114