Inquiry Into Income Convergence in MENA Countries:

A Neural Network Approach

Majid Sameti, Associate Professor of the University of Isfahan

E-mail: P_Sameti@Hotmail.com

Shekoofeh Farahmand, Assistant Professor of the University of Isfahan

E-mail: Farahmand_shr@Yahoo.com

Kayhan Koleyni, M.A of economic of the Islamic Azad University of Khorasgan (Isfahan)

E-mail: Kayhan_Koleyni@Yahoo.com

ABSTRACT

Economic convergence is one of the important topics of new macroeconomics. It refers to tendency of income per capita of countries (regions) to converge to their steady-state value. There are two kinds of convergence: conditional and absolute convergence. This paper examines income convergence between 22 MENA countries during the period of 1970-2003 by using the neoclassical growth model of Barro-Salla-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, the results become more precise. Results show that absolute convergence and conditional convergence are significant but the rate of convergence is low.

Key Words: *Income Convergence, MENA Countries, Artificial Neural Networks*
JEL: C_{45}, E_{37}, O_{47}

Introduction

Movement of the world toward integration, polarization and regional formation of blocks are current problems. Debates about integrations such as European, Islamic, G7 or ASEAN countries can be studied in different fields of economic emphasis. Income (output) convergence is one of the interesting problems of new macroeconomics. It refers to tendency of per capita income of countries (regions) to converge to their steady-state value.

Convergence hypothesis tries to answer two main questions. First, do the poor countries (regions) grow faster than rich ones? That depends on the effect of initial conditions on per capita income differences across countries and the speed of convergence, which introduced by β-convergence in growth literature. Second, one can ask whether the dispersion of per capita income of countries (regions) is decreasing over the time or not. This type of convergence, called σ-convergence, focuses less on initial conditions, and instead emphasizes income distribution by measuring standard deviations.

For 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, the conditional convergence is occurring (Durlauf et al., 2005). Expansion of literature on economic growth, its modeling, and the development of additional quantitative tools of analysis and different type of statistical information that can be used for quantitative analysis (cross-section, time series or panel data) have promoted a large body of empirical studies about convergence hypothesis. However, still some critics about both kinds of β and σ convergence exist. According to Durlauf et al. (2005), one of these criticisms about β-convergence is effects of linear approximation. There is a body of research that explores the effects of approximations that are employed to produce the models used to evaluate β-convergence. Durlauf and Johnson (1995), Binard and Pesaran (1999) Liu and Stengos (1999) represented evidence against the adequacy of linear approximation. On the other hand, Romer (2001) and Dowrick (2004) claimed that the approximation would be quite reliable. Accordingly, some of these studies show accuracy of linear approximation while the others not. In addition, Durlauf et al. (2005) declare that nonlinearity has deeper affect than simple approximation error and it can affect on steady state of per capita income and its identification problem, which is another criticism about β-convergence.

We try to solve these questions about non-linearity of the underlying relationships and ambiguity of functional form by using non-parametric approach of Artificial Neural Networks (ANN). We show that by changing the quantitative tools of analysis from traditional econometric tools to Artificial Neural Networks (ANN), the results will become more precise and the non-linearity problem will be solved and appropriate functional form of movements of per capita income of countries toward their steady-state value will be gained. We used Multilayered Feed-forward Networks for this purpose and compared it with OLS estimation method based on cross-country regression equations of Barro and Salla-i-Martin model (1990, 1992, and 2004) for both kinds of absolute and conditional β-convergence during the period of 1970-2003. This paper is organized as follows: in section 2 we discuss some main studies on convergence hypothesis, section 3 declares the methodology and ANN modeling. In section 4, we describe our findings and compare them with OLS method. Finally, section 5 is conclusions.

Literature Review

As mentioned before, a large number of studies according to type of data used, the countries in the question, the sample period in the question and choice of control variables exists but we describe here only little part of this huge body of empirics. So many of these empirical studies are based on neoclassical growth models such as Solow (1956), Swan (1956), Koopmans (1956), Cass (1965) and even the older work of Ramsey(1928). In addition, we can see some of studies with endogenous growth models

like Jones and Manuelli (1990) and Kelly (1992) but most of endogenous growth models are not compatible with convergence hypothesis like Romer (1986) and Lucas (1988) because of convexity in production function.

Convergence hypothesis originates from Abramovitz (1986) and Baumol (1986). Baumol used 1870-1974 data for 16 OECD countries and estimated the regression below. He focused the perfect convergence (b is approximately near with a value of -1). However later Delong (1988) showed that Baumol's conclusions are not correct because of the problems in sample selection and measurement error.

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

where, y is the income per capita 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 used the following cross-country regression:

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

Where, the subscript t denotes the year, and the subscript i denotes the country or region, y is income per capita, 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 the world. It is plausible that different regions within a country are more similar than different countries across the world with respect to technology and preferences. Then, as most of researches show, global absolute convergence does not exist. If the intercept a_{it} is the same for all regions 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 $(1 - e^{-\beta}).\log(\tilde{y}_i^*)$ 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, $\tilde{y}_i^*$. 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. We set steady state as follows:

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

Mankiw, Romer and Weil (1992) used augmented Solow model, which includes accumulation of human as well as physical capital for several large group of countries and found almost the same results as Barro and Sala-i-Martin (1992). Based on Barro

and Sala-i-Martin (1992) and cross-section approach, we then see a large body of research. There are also other approaches like time series and panel data analysis. Bernard and Durlauf (1995, 1996), Durlauf (1998), series of Quah's papers (1993a, 1993b, 1996a, 1996b, 1996c, 1997, 2001), Nahar and Inder (2002) used time series approach. This approach is largely statistical in nature, and disadvantage of this approach is that it is not according to particular modeled after growth theories.

Lee, Pesaran and Smith (1997), Islam (1995), Caselli, Esquivel and Lefrot (1996), Benhabib and Spiegel (1997), Nerlove (1996), Canova and Marcet (1995), Evans (1998) used panel data approach. The panel data analysis adapts the use of convergence equations as in cross-section analysis. Although, it can help to increase the flexibility of model, the structural error can be seen in this approach.

Papadas and Estratoglou (2004) tried to analyze β convergence by cross-section approach according to Barro and Salla-i-Martin model. They use the concept for 52 prefectures of Greek economy for two periods according to availability of investment data. They estimate both conditional and absolute convergence for period of 1981-1991 and 1971-1991. Additional variables for analyzing conditional convergence are percentage share of the total labor force employed in the primary sector, the percentage share of total population with secondary education, investment and unemployment rate. They introduce ANN algorithm as a useful tool for studying non-linearity relationship of β convergence. They utilize a Back-Propagation Network (BPN) with 10 neurons and 1 bias node and show it can substantially perform very well and more accurate. According to Papadas and Estratoglou (2004) there has been no other study of ANN application to the empirics of convergence and their study is the first. Albeit this fact, lack of analysis of an alternate neural networks is missing in their research. Efficiency of neural networks is extremely related to architecture and designing of these networks. Different networks with different architecture should be designed and among them, the best network with minimum error should be chosen. In addition, length of studied periods for neural networks is another discussing problem. Usually, neural networks with longer periods are more accurate.

Methodology

Artificial neural networks are the members of a family of statistical techniques, which try to simulate and model 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 neuron or node.

The difference between a neural network and other approximation methods is that the neural network makes use of one or more hidden layers, in which the input variables are transformed by a special function in parallel processing (McNelis, 2005). Each neuron has one ascendant activation function, which can be linear or nonlinear according to their

application. This activation function determine threshold of the neuron. The neuron receives a weighted sum of inputs from connected unit, and reply according to this threshold and weighted sum of inputs. Threshold behavior of *logsigmoid* and *tansig* or *tanh* activation function, which characterizes many types of economic responses to changes in fundamental variables, causes the great application of them in economy.

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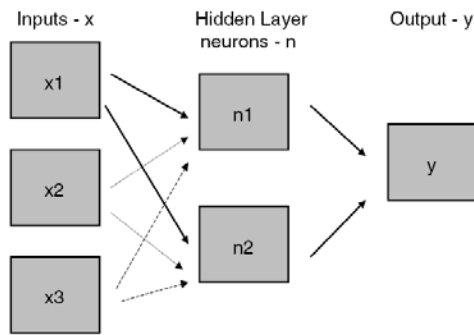
This figure shows the architecture of feed-forward back propagation networks. Inputs X makes the first layer of the networks. After this layer, hidden layer with n neurons processes the inputs in parallel. Final layer of a network is output layer.

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 (Schalkoff, 1997). 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 process exist: supervised learning and unsupervised learning (McNelis, 2005). 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 $i^{\bullet}$ input variables $\{x\}$, and $k^{\bullet}$ neurons. A linear combination of these input variables observed at time t, $\{x_{i,t}\}, i = 1,...,i^{\bullet}$, with the coefficient vector or set of input weights $\omega_{k,i}, i = 1,...,i^{\bullet}$, as well as the constant term $\omega_{k,o}$, 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 $k^{\bullet}$ neurons at time or observation index t are combined in a linear way with the coefficient vector $\{\gamma_k\}, = 1,...,k^{\bullet}$, and taken with constant term γ_o, 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
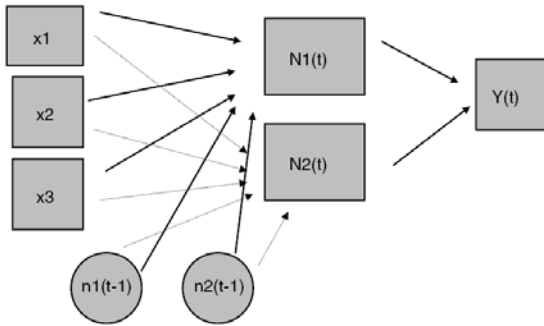
This figure shows the architecture of Elman Recurrent networks. Inputs X makes the first layer of the networks. After this layer, hidden layer with n neurons processes the inputs in parallel. Final layer of a network is output layer. This network has a feedback from the hidden layer that works like a 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:

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

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

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

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 a dynamic one 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 growth rates of real GDP per capita from 1970 to 2003 against levels of real GDP per capita in 1970. The straight line provides a best fit to the relation between the growth rate of real GDP per capita and the level of real GDP per capita. If the convergence prediction from Solow model were 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. As we can see, there is a very slight tendency for the growth rate to fall down with increase in the level of real GDP per capita. Most of the low levels income countries have high growth rates (more than 0.015). In addition, there is no country with negative growth rate. According to these facts, now we test convergence hypothesis with equations (2) and (3) from Barro and Sala-i-Martin model. Then ANN algorithm will be used to make β convergence estimation more accurate and automatically solve the nonlinearity problem of the definition. We used GDP per capita index from Penn World Table marked 6.2, which is the latest version of Summers and Heston's database (2006).

Figure 3: Growth Rate versus Level of Real GDP per Person for MENA Countries
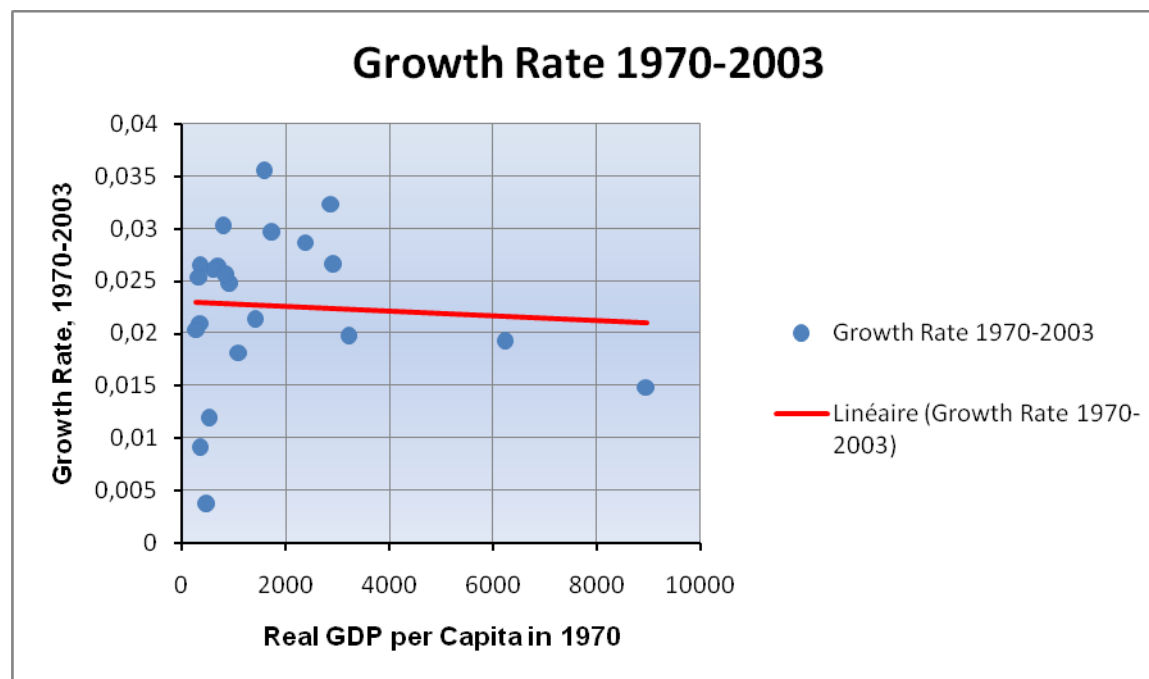
This figure shows MENA countries with different real GDP per capita in 1970 versus growth rate 1970-2003 on a proportionate scale. The red line is the trend line between countries data. If the slope of this trend line were negative means, that convergence exists. As we can see, the slop is almost negative which presents existence of convergence with low speed.

Table 1 shows the estimation results of equation (2) for 22 MENA countries[1] from 1970-2003. It includes the estimated value of the coefficient of independent variable with its *t-*

[1]-*Afghanistan-Algeria-Bahrain-Cyprus-Egypt-Iran-Iraq-Israel-Jordan-Kuwait-Mauritania-Morocco-Oman-Pakistan-Qatar-Saudia Arabia-Somalia-Sudan-Syria-Tunisia-Turkey-United Arab Emirates.*

value in parentheses, the corresponding derived value of β and the value of R^2. The estimated positive value of β, derived from the negative coefficient of log (y_{i,t_0}), which is statistically significant, demonstrates existence 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 initial conditions of 22 MENA 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
0.007	-0.007 (-4.94)	0.033

This table presents the results of linear regression model for absolute convergence, the positive sign of β support existence of absolute convergence with low speed.

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

$$\hat{g}y_{it} = 0.115 - \left(1 - e^{-0.007}\right) \cdot \log\left(y_{i,t-1}\right) \tag{10}$$

Accordingly, after we condition on steady state as described in equation (3) by adding different independent variables, we estimate the conditional convergence in Table 2. Variables added linearly to the original model include arable land (LND), life expectancy (EXP), annual growth rate of population (POP), openness in constant price (OPEN), investment share of real GDP (INV), metric tons per capita of CO_2 emissions (CO2) and percentage of government expenditure from GDP (GOV).

Table 2: Results of Conditional Convergence Regression

β	log(y_{i,t_0})	LND	EXP	POP	OPEN	INV	CO2	GOV	R^2
0.008	-0.008 (-5.45)	-5.87E-10 (-2.28)	----	----	----	----	----	----	0.039
0.007	-0.007 (-4.95)	----	-5.39E-06 (-0.074)	----	----	----	----	----	0.039
0.012	-0.012 (-6.62)	----	----	----	----	----	0.0006 (3.99)	----	0.058
0.008	-0.008 (-6.06)	----	----	0.0025 (4.24)	----	----	----	----	0.061
0.011	-0.011 (-7.08)	----	----	0.0022 (3.78)	0.00017 (3.66)	----	----	----	0.079
0.012	-0,012 (-7.67)	----	----	0.0024 (4.13)	0.00013 (2.92)	0.001 (4.71)	----	----	0.11
0.015	-0.015 (-7.94)	----	----	0.0017 (2.6)	0.00010 (2.23)	0.001 (4.94)	0.0004 (2.47)	----	0.118
0.012	-0.012 (-7.33)	----	----	0.0024 (4.09)	0.00014 (2.78)	0.001 (4.70)	----	-3.06E-05	0.107
0.013	-0.013 (-7.45)	----	----	0.0018 (2.41)	----	0.001 (5.18)	0.00046 (2.84)	0.00013 (0.93)	0.113
0.015	-0.015 (-7.41)	-5.24E-10 (-1.75)	2.61E-06 (0.037)	0.0020 (3.03)	6.86E-05 (1.26)	0.001 (5.13)	0.00033 (1.99)	-3.32E-05 (-0.22)	0.128

This table presents the results of linear regression model for conditional convergence, the positive sign of β support existence of absolute convergence. T-statistics in parentheses show 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 Table 2 coefficient of $\log(y_{i,t_0})$ is statistically significant with the negative sign. Additional variables like arable land, life expectancy, ratio of government expenditure to GDP are statistically insignificant and do not add to the model. Openness and CO2 emissions in some models are insignificant. Accordingly, the best model is:

$$\widehat{g}y_{it} = 0.124 - \left(1 - e^{-0.012}\right).\log\left(y_{i,t-1}\right) + 0.0024\,POP + 0.00013\,OPEN + 0.001\,INV \qquad (11)$$

We then use nonparametric ANN approach to make the estimation more accurate. A neural network uses three samples of data. Training sample that is presented to the network during learning and training process and the network is adjusted according to its error. Validation sample, which is used to measure network generalization and to halt training when generalization stops improving and testing sample that has no effect on training and so provide an independent measure of network performance during and after training. The most important challenge of neural network performance is related to its architecture. Usual way for 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% -70%, 20%,10% -80%, 10%, 10% and 60%, 20%, 20% which are more general orders for ANN algorithm and found 70%, 15%, 15% subsample is the best. Therefore, 1970-93 is the training sample, 1993-1998 is the validation sample and 1998-2003 is the testing sample. According to equations (4) to (9) we used BPN feed-forward network model as follow:

$$\log\left(\frac{y_{it}}{y_{i,t-1}}\right) = \gamma_0 + \sum_{k=1}^{k^*} \left[\gamma_k(\square) \frac{1}{1 + e^{-\left(\omega_{k,0} + \Sigma_{i=1}^{i^*}\left[\omega_{k,i}*\log\left(y_{it-1}\right)\right]\right)}}\right) \qquad (12)$$

And ELMAN recurrent network model:

$$\log\left(\frac{y_{it}}{y_{i,t-1}}\right) = \gamma_0 + \sum_{k=1}^{k^*} \left[\gamma_k(\square) \frac{1}{1 + e^{-n_{it}}}\right) \qquad (13)$$

$$n_{k,t} = \omega_{k,0} + \sum_{i=1}^{i^*} \omega_{k,i} * \log\left(y_{i,t-1}\right) + \sum_{k=1}^{k^*} \varnothing_k n_{k,t-1} \qquad (14)$$

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

Table 3: Results of Different Artificial Neural Networks for absolute convergence

Number of Network	Network Type	Number of Hidden Neurons	Percentage (Training-Validation-Testing)	MSE-TRAIN	MSE-TEST
1	BPN Feed-Forward	2	70%-15%-15%	0.00175	0.00225
2	BPN Feed-Forward	3	70%-15%-15%	0.00369	0.00444
3	BPN Feed-Forward	4	70%-15%-15%	0.00337	0.00308
4	BPN Feed-Forward	5	70%-15%-15%	0.000973	0.00281
5	BPN Feed-Forward	6	70%-15%-15%	0.000507	0.00904
6	BPN Feed-Forward	10	70%-15%-15%	0.0119	0.0168
7	BPN Feed-Forward	2	80%-10%-10%	0.00132	0.0373
8	BPN Feed-Forward	3	80%-10%-10%	0.00128	0.00893
9	BPN Feed-Forward	2	60%-20%-20%	0,00201	0.00396
10	BPN Feed-Forward	3	60%-20%-20%	0.00162	0.00725
11	BPN Feed-Forward	4	70%-20%-10%	0.00108	0.00336
12	ELMAN Recurrent Network	2	70%-15%-15%	0.00617	0.00301
13	ELMAN Recurrent Network	3	70%-15%-15%	0.00672	0.00467
14	ELMAN Recurrent Network	4	70%-15%-15%	0.00546	0.00735
15	ELMAN Recurrent Network	5	70%-15%-15%	0.00094	0.0347
16	ELMAN Recurrent Network	2	60%-20%-20%	0.0114	0.00864
17	ELMAN Recurrent Network	3	60%-20%-20%	0.00236	0.00311
18	ELMAN Recurrent Network	4	60%-20%-20%	0.00154	0.00475
19	ELMAN Recurrent Network	5	60%-20%-20%	0.00123	0.00349
20	ELMAN Recurrent Network	6	60%-20%-20%	0.00087	0.0121

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 networks are the first and fourth one.

As the table shows, different kinds of networks with different numbers of hidden neurons produce different performance. The best network with high performance will be chosen according to MSE of testing sample. The ability of the network to estimate accurately the annual rates of growth, which is our dependent variable, based on unused values of the independent variable in training indicates that it has sufficiently captured the underlying

relationship and the networks perform well with new data from testing sample, which are fed to networks after training. Despite the dynamic of Elman recurrent networks, the first and fourth neural network models BPN, feed-forward with two and five hidden neurons are those networks that minimize MSE in testing sample and can produce estimations that are more accurate. Table 4 presents the performance of these networks in total sample in comparison with linear regression model more specifically.

Table 4: Comparison of ANN and Linear Model Performance

| Model	MSE	MAE	MSEREG	SSE
BPN(2 neurons) Feed-Forward	0.0018	0.0282	0.0016	1.3083
BPN(5 neurons) Feed-Forward	0.0014	0.0252	0.0013	1.0381
Linear Regression	0.0021	0.0326	0.0019	1.5077

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

As the results show, mean squared error (MSE), mean absolute error (MAE), mean squared error with regularization (MSEREG) and sum squared error (SSE) of BPN feed-forward neural networks are less than those of linear regression model. Hence, neural network is a more capable and more accurate model than linear regression model. As the results show, ANN with five hidden neurons is the best one and is chosen.

We report values of R^2 for ANNs, but the definition of R^2 breaks down as we move away from traditional regression.

For the best vision about the ANN performance, Figure 4 shows in-sample evaluation of it. It illustrates a randomly selected year, 1990, which has been used during training in comparison with the real data. As can be observed, the network performs well in estimation of growth rates.
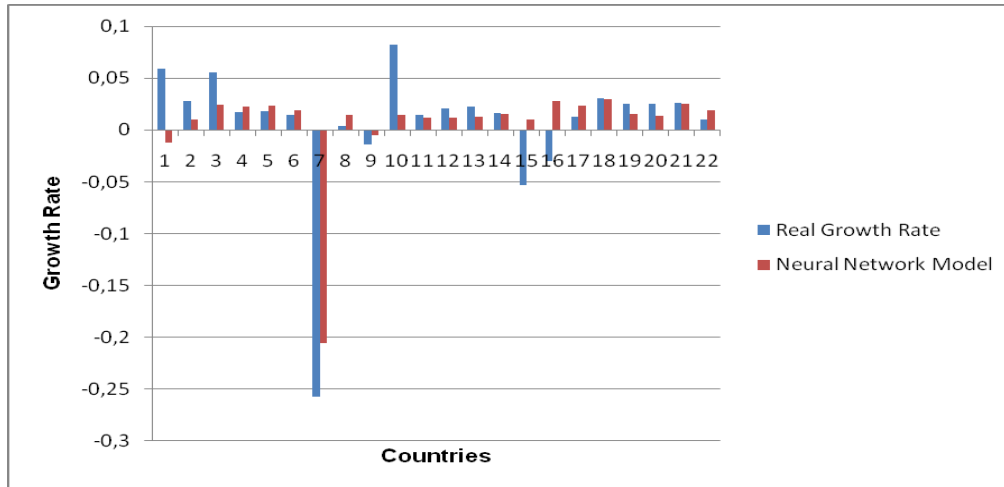
Figure 4: In-Sample Performance of ANN for 1990
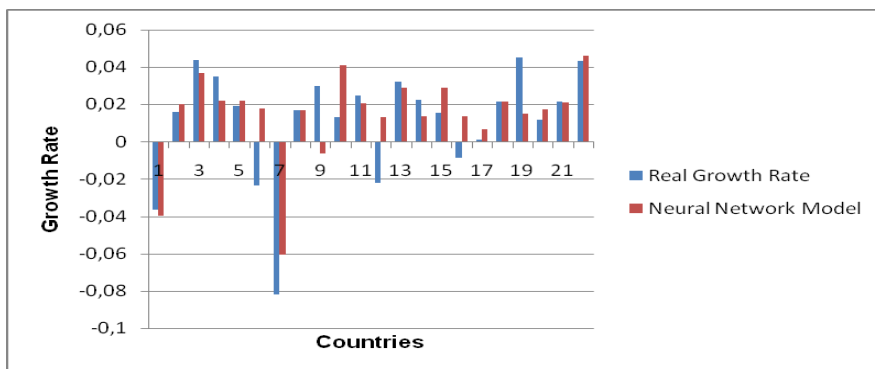
This figure shows In-Sample performance of artificial neural networks for 1990, randomly selected, for 22 MENA countries.

Figure 5 plots out-of-sample performance of ANN. It shows growth rates of 22 MENA countries in comparison with the estimated growth rates of selected neural network model for the year 2003, which is last period in testing sample. It is expected that estimation is less accurate than in-sample but still it is appropriate.

Figure 5: Out-of-Sample Performance of ANN for 2003

This figure illustrates Out-of-Sample performance of ANN. It compares the neural network estimations with real data of growth rate of MENA countries in 2003.

All of these figures and tables prove the capability of neural networks for studying income convergence and capturing the movement of different countries growth rates. Artificial neural networks also, can be used for analyzing unconditional income convergence. As Table 2 presents, additional variables like POP, INV and OPEN are statistically significant and we can use them in ANN models to study conditional convergence. Table 5 summarizes some of ANN models for conditional convergence.

Table 5: Results of Different Artificial Neural Networks for Conditional Convergence

Number of Network	Network Type	Number of Hidden Neurons	Percentage (Training-Validation-Testing)	MSE-TRAIN	MSE-TEST
1	BPN Feed-Forward	3	70%-15%-15%	0.00516	0.0108
2	BPN Feed-Forward	4	80%-10%-10%	0.00158	0.00372
3	BPN Feed-Forward	5	85%-10%-5%	0.000581	0.00226
4	ELMAN Recurrent Network	2	70%-15%-15%	0.00293	0.00268
5	ELMAN Recurrent Network	3	80%-10%-10%	0.00686	0.00573
6	ELMAN Recurrent Network	5	60%-20%20%	0.00115	0.00856

This table illustrates different Elman and BPN feed-forward networks with different performance and different errors for conditional convergence.

Model 3, with 5 hidden neurons is the best and more accurate network according to MSE of both testing and training samples. Like absolute convergence we can compare the performance of OLS and ANN models. Table 6 presents the results:

Table 6: Comparison of ANN and Linear Model Performance

Model	MSE	MAE	MSEREG	SSE
BPN(5 neurons) Feed-Forward	0.000933	0.021	0.000839	0.677
Linear Regression	0.0019	0.0319	0.0017	1.342

This table shows the performance of artificial neural networks and linear regression model with different measurement of their errors for conditional convergence. It supports that ANN is more accurate.
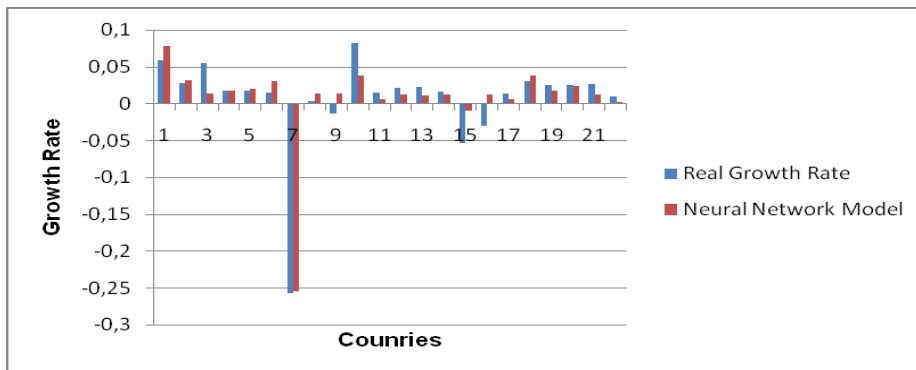
Again, the performance of neural network is better than linear regression model. Like unconditional convergence, in-sample and out-of-sample performance of 22 MENA countries are presented in Figure 6 and Figure 7.

Figure 6: In-Sample Performance of ANN for 1987
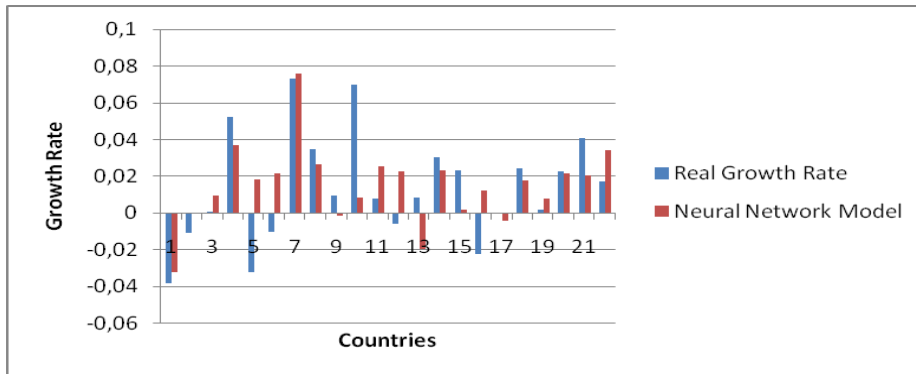
This figure shows In-Sample performance of artificial neural networks for 1990, randomly selected, for 22 MENA countries.

Figure 7: Out-of-Sample Performance of ANN for 2003

This figure illustrates Out-of-Sample performance of ANN. It compares the neural network estimations with real data of growth rate of MENA countries in 2003.

Again, like in absolute convergence, the results are accurate. Additional variables like OPEN, INV and POP make the out-of-sample performance of the model more accurate than in figure 5 but in both figures ANN performs so well.

Conclusions

As our results show, absolute convergence exists for 22 MENA countries across the world in studied period of 1970-2003. It means that our analysis supports tendency of poor economies to grow faster than rich ones across the MENA countries. In addition, after conditioning some different variables like openness, annual growth rate of population, investment, etc. we conclude that conditional β convergence is statistically significant in all of estimated models but the speed of convergence is low.

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 ANN algorithms. We show that by changing the quantitative tools of analysis and using ANN, the results become more precise in comparison with OLS method. For this purpose, we examine more than 50 neural networks with different architecture in two types of feed-forward back propagation (BPN) and Elman recurrent. We found despite the dynamic of Elman recurrent networks, the BPN, feed-forward neural network model with five hidden neurons is the one, which can produce estimations that are more accurate.

The most important point of this study is that although we compare neural network's accuracy with regression model and conclude neural networks are more capable, we used artificial neural networks as the complement algorithm of OLS method, not as the alternative or substitute approach.

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

References

Abramovitz, M. (1986). "Catching up, forging ahead and falling behind". *Journal of Economic History* 46, 385–406.
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.

Baumol,W. (1986). "Productivity growth, convergence, and welfare:What the long-run data show". *American Economic Review* 76 (5), 1072–1085.

Benhabib, J. and Spiegel, M. M., (1997), "Cross-Country Growth Regressions",

Working Paper 97-20, CV Starr Center, New York University.

Bernard, A., Durlauf, S. (1995). "Convergence in international output". *Journal of Applied Econometrics* 10 (2), 97–108.

Bernard, A., Durlauf, S. (1996). "Interpreting tests of the convergence hypothesis". *Journal of Econometrics* 71 (1–2), 161–173.

Canova, F. and Marcet, A., (1995), "The Poor Stay Poor: Non-Convergence Across Countries and Regions", Discussion Paper 1265, CERP.

Cass, D. (1965). "Optimum Growth in an Aggregative Model of Capital Accumulation". Review of Economic Studies (Blackwell Publishing) 37 (3): 233–240. http://www.jstor.org/stable/2295827. Retrieved on 2008-04-17.

Caselli, F., Esquivel, G., Lefort, F. (1996). "Reopening the convergence debate: a new look at cross country growth empirics". *Journal of Economic Growth* 1 (3), 363–389.

DeLong, J.B. (1988). "Productivity growth, convergence, and welfare: Comment". *American Economic Review* 78 (5), 1138–1154.

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.

Evans, P. (1998). "Using panel data to evaluate growth theories". *International Economic Review* 39 (2), 295–306.

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. (1995). "Growth empirics: A panel data approach". *Quarterly Journal of Economics* 110 (4), 1127–1170.

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

Jones, L., Manuelli, R. (1990). "A convex model of equilibrium growth: Theory and policy implications". *Journal of Political Economy* 98 (5), 1008–1038.

Kelly, M. (1992). "On endogenous growth with productivity shocks". *Journal of Monetary Economics* 30 (1), 47–56.

Koopmans, T.C. (1956). "The Klein-Goldberger Forecasts for 1951, 1952 and 1954, Compared with Naive-Model Forecasts," Cowles Foundation Discussion Papers 12, Cowles Foundation, Yale University.

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.

Lucas, R. (1988). "On the mechanics of economic development". *Journal of Monetary Economics* 22 (1), 3–42.

Mankiw, N.G., Romer, D.,Weil, D.N. (1992). "A contribution to the empirics of economic growth". *Quarterly Journal of Economics* 107 (2), 407–437.

McNelis, P.D., (2005). "*Neural Networks in Finance: Gaining Predictive Edge in the Market."* Elsevier Academic Press. 118-224.

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

Nerlove, M. (1996). "Properties of alternative estimators of dynamic panel models: An empirical analysis of cross-country data for the study of economic growth". In: Hsiao, C. (Ed.), Analysis of Panels and Limited Dependent Variable Models: In Honor of G.S. Maddala. Cambridge University Press, Cambridge.

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. (1993a). "Galton's fallacy and tests of the convergence hypothesis". *Scandinavian Journal of Economics* 95, 427–443.

Quah, D. (1993b). "Empirical cross-section dynamics in economic growth". *European Economic Review* 37 (2–3), 426–434.

Quah, D. (1996a). "Twin peaks: growth and convergence in models of distribution dynamics". *Economic Journal*106 (437), 1045–1055.

Quah, D. (1996b). "Empirics for economic growth and convergence". *European Economic Review* 40 (6), 1353–1375.

Quah, D. (1996c). "Convergence empirics across economies with (some) capital mobility". *Journal of Economic Growth* 1 (1), 95–124.

Quah, D. (1997). "Empirics for growth and distribution: Stratification, polarization, and convergence clubs". *Journal of Economic Growth* 2 (1), 27–59.

Quah, D. (2001). "Searching for prosperity: A comment". Carnegie-Rochester Conference Series on Public Policy 55, 305–319.

Quah, D. and Durlauf, S. N., (2004). "Handbook of Macroeconomics" 4 (2), 238- 256.

Ramsey F.P. (1928). "A mathematical theory of saving". *Economic Journal*, vol. 38, no. 152, 1928, pages 543–559.

Romer, P. (1986). "Increasing returns and long-run growth". *Journal of Political Economy* 94 (5), 1002–1037.

Schalkoff, R. J. (1997). "*Artificial Neural Networks*". Mcgraw-Hill Press.25-44.

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.

Swan, T. (1956). "Economic growth and capital accumulation". *Economic Record* 32, 334–361.

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**.**