

Papadas, Christos T.; Efstratoglou, Sophia

**Conference Paper**

## Estimation of Regional Economic Convergence Equations Using Artificial Neural Networks with Cross Section Data

44th Congress of the European Regional Science Association: "Regions and Fiscal Federalism", 25th - 29th August 2004, Porto, Portugal

**Provided in Cooperation with:**

European Regional Science Association (ERSA)

*Suggested Citation:* Papadas, Christos T.; Efstratoglou, Sophia (2004) : Estimation of Regional Economic Convergence Equations Using Artificial Neural Networks with Cross Section Data, 44th Congress of the European Regional Science Association: "Regions and Fiscal Federalism", 25th - 29th August 2004, Porto, Portugal, European Regional Science Association (ERSA), Louvain-la-Neuve

This Version is available at:

<https://hdl.handle.net/10419/117019>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Estimation of Regional Economic Convergence Equations Using Artificial Neural Networks with Cross Section Data

Christos T. Papadas and Sophia Efstratoglou \*

## *Abstract*

*Theoretical developments and discussions on growth and regional convergence have been accompanied by another debate, associated with the type of data used and quantitative approaches adopted in empirical research. Estimation of convergence equations continues to play a key role in the study of economic convergence, despite criticisms. This paper introduces the use of artificial neural networks (ANN) in the study of convergence. It focuses on the concept of  $\beta$ -convergence and accepts that cross section data can provide useful information for its investigation. Non-linearities of the underlying relationships, the restrictiveness of assumptions on functional forms, and econometric problems in the estimation and application of certain theoretical models, advocate for the use of ANN algorithms. A back-propagation (BPN) artificial neural network is constructed and utilized to study convergence of regional, gross domestic products per capita in Greece, together with the application of a traditional econometric analysis. Cross-section statistical data on Greek prefectures are used while results and repeated testing show that the neural network performs very well in estimating convergence equations. It improves substantially the accuracy of estimates and predictability of the estimated relationships. In addition, the BPN algorithm could be used with time series or panel data, and it could estimate also convergence equations of additional economic or social variables.*

\* The authors are Lecturer and Professor respectively, at the Department of Agricultural Economics and Rural Development, at the Agricultural University of Athens, Greece.

## **1. Introduction**

An expansion of the literature on economic growth and its modeling, respective implications for regional convergence, and the development of additional quantitative tools of analysis, have promoted the debate over convergence issues. Part of this debate relates to the choice of method of quantitative analysis. The relevant decision made, is most often dependent on the adoption of a theoretical model analyzing the process of growth and convergence. It is also dependent on the type of statistical information that should be used for quantitative analysis (cross-section, time series or panel data).

As it is so often stated, two basic concepts of economic convergence exist in recent literature. It is argued that  $\beta$ -convergence exists when regions or countries with lower initial per capita incomes grow faster than regions or countries with initially higher per capita incomes. In addition,  $\sigma$ -convergence exists when the dispersion of the per capita income distribution, as measured by its standard deviation, declines. The first concept of convergence is a necessary but not sufficient condition for the second. Hence, while  $\sigma$ -convergence implies  $\beta$ -convergence, the opposite is not true.

The two concepts of convergence, as used and quantified in the literature, are far from perfect. In the case of  $\sigma$ -convergence for example, there may be an increase or decrease in the dispersion around the mean of the distribution, but the situation may be different within various groups of observations. Therefore, the overall picture of a change in the dispersion, as measured by the standard deviation, may not be enough to demonstrate if there has been a general trend to reduce or increase the gap in regional per capita incomes. On the other hand, using the concept of  $\beta$ -convergence is not without problems either. The testing for  $\beta$ -convergence and its quantification ("speed" or "rate" of convergence) relies on the estimation of convergence equations. These equations are derived from models of economic growth. The form of these equations and the meaning of certain of their parameters may differ, according to the assumptions of the models and the derivation process of the convergence equations. Hence, altering the assumptions of the underlying growth models and the derivation process can put into question the specification of convergence equations.

The proposition and adoption of different theoretical growth models with different implications or estimations for  $\beta$ -convergence, the inability of the latter to tell the whole story on convergence, and the inadequacies of  $\sigma$ -convergence as well, have led to criticisms and more recently, to few specific propositions of new methodologies to capture and measure convergence without using the two concepts. Generally speaking, these methodologies are based on the analysis of changes in the overall distribution of regional data. Nevertheless, new methodologies avoiding the use of the two concepts are not without problems either. Not only it is questionable if they are truly more informative, but they rely also on strict assumptions that the data set must satisfy.

It is not strange therefore that most of the quantitative analyses on regional convergence utilize the two concepts and analytical discussions and methodological developments are focusing on these two concepts. The necessity of estimating convergence equations and their significance, are maintained, and this is reflected on a growing part of the literature on regional convergence. Given the problems and the debate associated with convergence equations, we believe that artificial neural networks (ANN) offer an alternative approach to their estimation. They can contribute substantially to facing data and computational problems and improve the accuracy and predictability of the estimated relationships.

The major objective of this study is to investigate regional economic convergence in Greece implementing and evaluating also the performance of artificial neural networks. For this reason, a feed-forward ANN of the back-propagation form (BPN) is constructed and proposed for the estimation of convergence equations. The BPN algorithm is used with cross sectional data on per capita gross domestic products and other variables for the Greek prefectures and for three census years. A traditional econometric approach is also used to examine  $\beta$  - convergence together with the simpler testing for  $\sigma$ -convergence. The role of conditioning variables such as the level of private investment, education, etc, is also investigated. The significance and role of these variables is examined when they are considered either at the initial state or as flows and changes during the whole period examined.

## **2. A Review and Discussion on the Literature**

The original neoclassical growth models of closed economies proposed by Solow (1956), Swan (1956), Koopmans (1965), Cass (1965) and even the older work of Ramsey (1928), placed assumptions on the production function and technology some of which varied while others did not. Such assumptions relate to concavity and the degree of homogeneity, the existence or not of increasing returns, the treatment of endogeneity or not of savings, the issue of free access to an exogenous technology, and the way that human capital is incorporated in the analysis. There is some flexibility in the results with regards to output growth sustainability. In general however, these early closed economy models did imply convergence and the rate of per capita income growth over a period is inversely related to the initial per capita income.

Based on similar assumptions and endogenously determined savings, this convergence result ( $\beta$ -convergence), is upheld in the more recent, seminal work of Barro and Sala-i-Martin (1992, 1995). The original neoclassical models were more flexible than what is usually acknowledged by critics. Violations of the diminishing returns assumption as well as economic divergence, were not excluded altogether. In the case of Solow (1956) for example, long run income levels depend on rates of investment and population increases and there are possibilities of not achieving always the results on growth and convergence. Diffusion and access even to an exogenously available technology depend on several critical factors often overlooked or treated as exogenous when in fact they are not Abramovitz (1979, 1986). Deviations from the original results on growth and convergence due to such factors are possible.

In addition, different than the Barro and Sala-i-Martin assumptions in the newer neoclassical modeling can lead to different results on convergence. In Romer (1986) it is shown how the assumption of increasing input returns can invalidate the original conclusions of diminishing rates of growth and  $\beta$ -convergence. Increasing returns are often related to endogenous technological change. Even in the case of such a change however, the diminishing rates of growth and  $\beta$ -convergence can still hold as long as the endogenous technology remains a pure public good. The same holds also if the

new technology becomes available to all regions in the long-run (de la Fuente 1997). Endogenous technological change is frequently related to human capital improvement but this depends on how human capital is modeled and incorporated in the production function (Lucas, 1988, Grossman and Helpman, 1991, Romer, 1999, Mankiew, Romer and Weil, 1992). Different functional forms can be adopted to capture among other things the possibility of increasing returns (e.g. Rebelo, 1991, Duffy and Papageorgiou, 1997, Jones and Manuelli, 1990, Durlauf, 1993, 1996, etc.).

As a consequence, the result of economic convergence in the traditional neoclassical models and in those with endogenous technological change is a conditional situation. Convergence systematically derived as a result of these models is simply a trend towards stabilization of relative incomes, while the long run equilibria can be very different due to different conditions. Since this concept of convergence is not incompatible with geographical inequalities, it cannot be examined empirically using the traditionally available indices of dispersion and unequal distribution, even though the latter maintain their interest in the course of economic growth. These realizations reflect the reason for adopting both of the two distinct concepts of convergence.

The form of convergence equations most widely used in the cross section study of  $\beta$ -was derived by Barro and Sala-i-Martin (1990, 1992). The dependent variable is the average rate of growth of per capita income during the considered period, while the per capita income at the initial state is an independent variable. More specifically, a production function with output per unit of effective labor as the dependent variable and capital per unit of effective labor as the independent variable, which satisfies the usual convexity and continuity properties is assumed, while effective labor is a function of the exogenous technological development and an increasing function of the labor (population) size. The economy is closed and the capital per unit of effective labor changes over time according to a specified differential equation, which includes the production function and the capital per unit of effective labor itself, while the depreciation rate, the growth rate of labor (population) and the level of consumption per unit of effective labor enter as parameters.

The model assumes a representative household with preferences and a certain utility function with respect to consumption per person, such that the marginal utility has a constant elasticity with respect to this consumption. In the dynamic maximization of utility problem, the model assumes also another simple relationship between the parameters of the production and utility functions, in order to have the transversality condition satisfied. Then, the first order conditions are derived and so is the relationship to be satisfied by the steady state of capital per unit of employed labor. At the steady state, the per unit of effective labor variables of output, capital and consumption, remain constant and the per capita quantities of output, capital and consumption, increase at a rate equal to the rate of the exogenous, labor augmenting technological progress. If the initial value of capital per unit of effective labor is less than its steady state, it approaches to the latter monotonically with respect to time. It has been shown also (Barro and Sala-i-Martin, 1991) that the rate of increase of capital per unit of labor (per capita or per worker), is being reduced monotonically with respect to time towards its steady state rate.

If it is assumed that the production function is of the Cobb-Douglas form (with the exponent values between zero and one), the same result as above can be extended to the rate of increase in the output per unit of labor. The consequence of that is that if two economies have the same utility, and production technology, the economy with the lower initial capital per unit of effective labor will tend to grow at higher rates than the other economy.

#### Derivation of convergence equations – Cross Section Analysis

In the Barro and Sala-i-Martin model, under a Cobb-Douglas production function, the dynamic process described and discussed above is derived and expressed from the first order conditions of the utility maximization problem and the differential equation describing the motion of capital per unit of effective labor, over time. If the derived expression of the process is subjected to a logarithmic transformation and linearization, and modified from continuous to discrete time, after some operations it becomes:

$$\log\left(\frac{y_{it}}{y_{i,t-1}}\right) = a_i - (1 - e^{-\beta})[\log(y_{i,t-1}) - x_i(t-1)] + u_{it} \quad (1)$$

where  $a_i = x_i + (1 - e^{-\beta}) \log(\tilde{y}_i^*)$  and  $u_{it}$  is the disturbance term. Here,  $t$  denotes time,  $y_{it}$  is the per capita income of region  $i$  at time  $t$ ,  $(\tilde{y}_i^*)$  is the steady state of the per capita income per unit of effective labor in region  $i$ , while  $\chi_i$  is the steady state rate of increase of capital per unit of labor. In addition,  $\beta$  is a function of the preference and production parameters, of the steady state rate mentioned above, and of growth and capital depreciation rates. This parameter is an expression of the speed of adjustment towards the steady state of per capita income (rate of convergence). Obviously, a common value of  $\beta$  for all regions  $i$  is assumed here. This may not be the case always but it's much closer to reality when dealing with same country regions, which are likely to present similarities in terms of the parameters involved. Moreover, technology differences due to available methods of production or natural resources, which are reflected in the value of the constant multiplier in a Cobb-Douglas production function, do not affect the value of  $\beta$  anyway (King and Rebello, 1989).

If the relationship that describes the dynamic process in Barro and Sala-i-Martin has both sides divided by the number  $T$  of years in the period examined to get average values then, using (1) under the assumption that the steady states and steady state rates  $(\tilde{y}_i^*)$  and  $\chi_i$  respectively, are the same in all regions  $i$  ( $\tilde{y}_i^* = \tilde{y}^*$  and  $x_i = x$ ), then it is shown that:

$$\frac{1}{T} \log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = B - \left(\frac{1 - e^{-\beta T}}{T}\right) \log(y_{i,t_0}) + u_{i,t_0,t_0+T} \quad (2)$$

where the last term in (2) is the distributed time lag of the errors  $u_{it}$  between time  $t_0$  and  $t_0+T$  where  $B = x + [(1 - e^{-\beta T})/T][\log(\tilde{y}^*) + xt_0]$ . The dependent variable is now the annual average rate of growth over the whole period. Shifts in the constant term are due to the technology trend with a change in the initial time  $t_0$ . An implication of the discussion so far is that when  $\beta > 0$  we can argue that we have  $\beta$ -convergence. When  $\beta < 0$  there is no  $\beta$ -convergence and regions with greater initial per capita income tend to grow faster than the initially poorer ones. In this case there can be no  $\sigma$ -convergence either since  $\beta$ -convergence is a necessary condition for that. The estimation of the convergence equation in (2) or one of its variants is the most widely



used approach for the testing and measurement of  $\beta$ -convergence using cross-section data.

The econometric model is used either as implied here, or augmented (“constrained”) with additional conditioning variables, explanatory of the average rate of growth. The significance of such conditioning variables indicates in the case of  $\beta$ -convergence that convergence is “conditional”, in which case steady states to which different regions converge are different. If  $\beta > 0$  and  $\beta$ -convergence is not conditional (“absolute”), per capita incomes tend towards the same steady state in spite of any possible short-run exogenous disturbances in their trajectory. Conditioning variables usually relate to the level of investment, population, employment in the primary and non-primary economic sectors, institutional factors, available infrastructure, etc. With the exception of some flow variables, most of the conditioning variables are taken at their initial state values but this is not necessary, since their changes over the whole period have been used in the literature as well.

Several comments are in order here, with regards to the use of cross-section data and the form of convergence equation adopted. The appeal of cross-section data analysis is based not only on theoretical grounds such as the modeling discussed, but also on common economic logic. If poorer regions tend to grow faster than richer ones, then for a set of regions with different levels of per capita income at the beginning of the period, a negative relationship between the average rate of growth over the period and the initial per capita income should be expected and found. The criticisms we discuss with regards to cross-sectional analysis, certainly do not refute this argument.

The empirics of growth and convergence do present substantial similarities for a variety of theoretical models with different assumptions and implications. The assumptions of Barro and Sala-i-Martin for example support the expectation of  $\beta$ -convergence which is not necessarily the case with the results of Mankiw, Romer and Weil. However, these approaches result in similar convergence equations. Of course, the constant term and the slope are different expressions of the same and different parameters included in the models, and even the  $\beta$  coefficient could be interpreted differently. Nevertheless, the same form of convergence equation derived is also

estimated empirically, regardless of the concepts and theoretical arguments underlying its parameters. The empirical estimation allows the testing of theoretical results such as the existence of  $\beta$ -convergence.

It is both, the argument on economic logic and the theoretical modeling mentioned that has justified using convergence equations of the form given in (2) especially in cross-section analysis. An argument against this type of analysis is that it does not consider and provide information on the dynamics of the convergence process. Results rely on observations at the beginning and the end of a period. Convergence is a dynamic process and cross-section analysis on the same regions, over different periods for different per capita income ranges, may lead to different results. Depending on the complexity of the dynamics, differences may refer to the conclusion on  $\beta$ -convergence altogether, or the speed of convergence as well.

#### *Time-Series and Panel Data Analysis*

Different approaches based on time series analysis have been proposed (Bernard and Durlauf 1995, 1996), Durlauf (1989), Quah (1992), etc. Convergence in these approaches is investigated as a characteristic of the relationship between long-run forecasts of per capita outputs of different regions, given certain initial conditions. It leads therefore to a different type of convergence equations than the mentioned cross-section analysis where convergence is investigated as a characteristic of the relationship between initial income and average growth rate over a certain period.

In Bernard and Durlauf [1996], time series forecast convergence between two economies is defined as the equality of long-term forecasts of per capita incomes at a given date, and for given information available at this date. It can be shown easily that time series forecast convergence implies  $\beta$ -convergence when growth rates are measured between  $t$  and  $t+T$  for some fixed finite horizon  $T$ . The main conceptual difference between the two approaches is that an expected reduction in contemporary differences as implied by  $\beta$ -convergence differs from their eventual disappearance as implied by the time-series forecast convergence (Durlauf and Quah, 1998).

In empirical time-series analysis, it can be tested whether the differences between per capita incomes in selected pairs of regional economies, are described by a zero-mean stationary stochastic process and forecast convergence can be tested using unit root and cointegration analysis. Non-zero (i.e. deterministic) time trends in the cross-pair differences of per capita incomes or the presence of a unit root lead to rejection of a convergence hypothesis. The method is followed as a result of the recognition that convergence between pairs of economies, does not imply convergence for all economies, unless of course convergence holds for all pairs.

Two main approaches applying the discussed conceptual framework of time-series analysis discussed, are given by Bernard and Durlauf (1995, 1996) and by Quah (1992). In the first, analysis is confined to smaller groups of certain economies, which allows the use of long time-series data. Unit root tests and cointegration analysis can again be applied to derive conclusions on convergence for the specific group under study (the authors used data on the OECD countries). However, between pairs of individual economies in the group, the situation may still be different than the derived conclusion for the whole group. In the second approach proposed, what is studied is the existence of common stochastic trends in a large group of economies. In particular, the differences in per capita outputs between each economy in the group and a given chosen economy are utilized (in Quah's study in particular, the per capita output differences between a large group of countries and the US are taken).

It is clear that seeking to extract empirical information on the dynamics of growth and converge does not suffice on its own to render invalid the results of cross-section analysis especially if the purpose of the latter is confined to answers on the existence of convergence and its measurement over a period. The previously mentioned criticism related to convergence dynamics and possibly changing results over different periods, would be more difficult to sustain when cross-section analysis covers relatively large periods (especially several of them). In addition, it would be a mistake to consider that other approaches such as time-series analyses are freed from such problems. The period covered by the time series data and the specification of the model adopted do affect the conclusions of the analysis on convergence. In fact, for a generalized conclusion there must always be the heroic assumption, that an estimated

chosen form of a difference equation describes a process well outside the period covered by available data, but if the process is complex enough the conclusions may be misleading again. The issue of convergence or not for a group of regions, versus the behavior of sub-groups or individual regions, has not been resolved satisfactorily either and it is impossible that a complete and satisfactory answer can be given within the framework of any convergence equation estimation.

The various versions of time-series analysis suffer in general from serious specification and other econometric problems that create strong doubts for the accuracy of their results and the reliability of their conclusions. It has been argued that a major disadvantage of time series analysis is that it assumes that the data used are described by a time invariant data generating process, which is not satisfied by per capita output and income series if the respective economies are moving towards their steady state (Durlauf and Quah, 1998). From the start, this creates a conflict with cross-section analysis since the requirement for convergence that output or income differences are stationary and zero-mean, is inconsistent with the requirement that the difference between a richer and a poorer economy has a non-zero mean, implied in cross-section empirical estimations (Bernard and Durlauf, 1996).

However, time-series approaches can differ substantially between themselves and lead to different results using the same data set. Nahar and Inder (2002) propose a time series approach to identify particular countries within a group that may not follow the overall tendency towards convergence. The authors dispute also some of the most prominent time-series approaches arguing that standard unit root or cointegration tests such as those adopted in the procedure of Johansen (1988, 1991), Bernard and Durlauf (1995) are inappropriate. Discussing the empirical convergence requirement for zero-mean stationarity, they show how this can create problems since there are certain non-stationary processes that meet their definition for convergence and a stationarity test may easily accept the unit root hypothesis concluding that there is no convergence. This may explain why the authors found evidence of convergence in OECD countries using their time series approach, while Bernard and Durlauf (1995) and Quah (1992) had found the opposite, in contrast to cross-section analysis. On similar grounds the authors criticize the panel data approach of Evans and Karras

(1996), based on their own alternative version and definition of convergence, which again is not equivalent to the empirical requirement of stationarity. Evans and Karras had criticized traditional OLS and NLS econometrics to estimate convergence equations, arguing that it was resulting in correlation between the error term and right hand side variables but the criticism is not confined to cross-analysis, even though their own proposed procedure did not substantially alter empirical results in their study. The authors found in favor of conditional convergence among the US states and a large group of countries.

There are certain limitations in time series analysis with regards to the use of conditioning variables in order to capture the significance of different characteristics though convergence itself can be examined as absolute or conditional. There is also the persistent problem of ignoring other, unobservable or not easily measured economy specific characteristics. It is some of the efforts to face this problem that advocated for the use of paned data methods to study cross country data on per capita incomes, with prominent approaches in terms of applications those of Lee, Pesaran, and Smith (1997) and Islam (1995). Additional contributions can be found in Caselli, Esquivel, and Lefort (1996), Benhabib and Siegel (1997), Nerlove (1996). Within the traditional econometric analysis of panel data, cross country unobserved heterogeneities (“individual effects”) are treated as nuisance parameters to be removed (exceptions and different approaches are found in Levin and Lin (1992), Quah (1994), Canova and Marcet (1993) and Evans (1998).

The panel data analysis adopts the use and estimation of convergence equations as in the cross-section analysis and of the form presented above. A general discussion on the derivation of panel data convergence equations based on the Mankiw, Romer and Weil model can be found in Durlauf and Quah (1998). The two main differences now are, first that the changes in per capita incomes of regions or countries in each observation are considered between successive periods  $t$  and  $t+1$  (successive period adjustments), and the second that the constant term is decomposed into time specific and economy specific effects.

It has been argued that the decomposition of the constant term adds to the appropriateness of the panel data analysis to study growth dynamics. Islam (1995) in particular argues that time and country effects are more probable to occur when the convergence equation is used to study convergence of per capita outputs than effective labor productivity. In addition, the error structure can be seen as a consequence of omitting variables from the equation and the two effects into which the constant term is decomposed can have several alternative interpretations. It is argued that the above increase the flexibility of the model and should reduce the possibilities for misspecification. (Durlauf and Quah,1998).

However, as Durlauf and Quah discuss also, there is a possibility that the decomposition of the constant term leads to misleading results or their misinterpretation with regards to convergence between poorer and richer regions. The circumstances that this can occur are not exceptional. He also points out to the usual econometric problem of correlation between country or region specific effects country and other right hand variables in the Mankiew, Romer and Weil version of the convergence equation on which the panel data analysis is usually based (the same could hold for conditioning variables in the Barro and Sala-i-Martin case).

Solutions to this problem, which affects the consistency of estimations are far from perfect and can lead again to misinterpretations of results (Durlauf and Quah in particular, argue that the problem is more serious “profoundly limiting” our ability to explain patterns of cross-country growth and convergence). There are still also differing views on the time period that the application of a growth model (i.e. the panel data analysis now) should cover. This is according to Durlauf and Quah the dual of the problem created after the conditioning out the country specific effects, which relates to the fact that only high frequency income movements are left to be explained, a problem related to the possibility of deriving misleading results). He argues in favor of long-run analysis but Islam (1995) points out that the convergence equation in of Mankiew, Romer, and Weil (the same could be said for the version of Barro and Sala-i-Martin and others as well) on which the panel data empirical models are based, are derived as we saw in the Barro and Sala-i-Martin case, using a log linear approximation around the steady state. Therefore, panel data analysis presents

more validity over shorter periods in time. The argument is rational but not convincing for all (e.g. Quah, 1998), especially if one cares more for the empirics and less for the theoretical foundation and consistency of the model used.

There have been several attempts to study convergence using "clustering and classification" (Durlauf and Johnson, 1995). Cross-section estimates of convergence equations are derived for the subsamples of a given data set. The subsamples are the product of splitting the data set on regions or countries into different initial per capita income classes. Using the data subsample of every income class, a different cross-section convergence equation is estimated for every class (Durlauf and Johnson used as criteria for their classification of observations not only per capita incomes but also literacy rates). The identification of threshold levels that define the different classes was done endogenously using regression-tree procedures. Results display significant differences between the different classes. Naturally, there are other ways and criteria too to conduct the classification and then estimate the convergence equations (Franses and Hobijn, 1995).

Since the concept of  $\beta$ -convergence is perceived as not saying the whole story and other measures such as  $\sigma$ -convergence are not sufficient either, few of the more recent efforts have focused on developing ways to analyze changes in the data (e.g. per capita income) distribution, without resorting to convergence equation estimation (either using cross-section, time-series or panel data), or without using single parameters of the distribution. These methods have not found as many applications yet as the empirical approaches we referred to. An approach which has been followed in some studies can be found in Quah (1993c). Quah uses Markov processes and ergodic theory to study convergence. His procedure requires the construction of a matrix resembling to, and treated as a transition probability matrix. Each row and column represent classes of per capita income and each cell is assigned the total number of observation units (i.e. countries or regions) that have transferred from one income class to another. The total sum of transitions for all time units we have data for is considered, within the examined period. Conclusions on convergence are derived using the characteristic roots of that matrix. An obvious problem of this approach is the arbitrary nature of income discretization in defining the income

classes. Different discretization may well lead to different results. A serious additional problem is the assumption that the data follow a stationary first-order Markov process.

Bickenbach and Bode (2003) proposed Pearson,  $\chi^2$ , and likelihood ratio tests to examine the validity of the above hypothesis. They applied their test to a data set of per capita incomes for the US states during the period 1929-2000 and they found that data do not follow a stationary first-order Markov process. It is clear that the results of this method can be put to question. There have been other attempts and ways, less widely known and adopted to study convergence by looking at changes of the whole distribution while facing the problem of the arbitrary discretization. Such a case is the adoption and use of stochastic kernels (Bianchi 1997, Jones (1997), Quah, (1997, 1996). Some of these approaches have similarities in their logic with the justification of using artificial neural networks as well. However, as it is always the case, regardless the progress we make in developing techniques to study changes in distributions, there are no results, parameters, etc. which can always tell the whole story as well as tediously examining each observation with regards to the others and the relevant changes.

These techniques, which focus on distributional behavior and changes, are not based or derived from any solid theoretical background. They do little to explain the confirmation or rejection of a theoretical conclusion. For this reason, we think it's an irony and unfair that Durlauf and Quah (1998) criticize Barro (1997) for pointing out that empirical analysis based on theoretical models with different assumptions than the Barro and Sala-i-Martin model, has proved again the validity of their own theoretical results. The criticism was based on the notion that the understanding of a phenomenon should derive from a better than its alternatives theoretical model, while compatibility of empirical results and observations with a model's conclusions does not suffice.

Despite the usefulness for the empirics of convergence that methods studying the distributional dynamics may have, despite also the limitations in information provided by the existence or not of  $\beta$ -convergence, the estimation of convergence equations can



play a crucial role for our understanding of the processes and factors involved. The existence of a relationship between the initial income per capita and rates of growth for example, needs to be investigated and it maintains always its theoretical interest regardless of achieving or not convergence at the end. Empirical approaches to convergence that avoid the estimation of convergence equations conceal that relationship, they suffer from serious flaws, and they can offer no help in studying the possible role of conditioning variables affecting convergence and its speed.

We have chosen to conduct a cross-section analysis of regional convergence in Greece- estimating convergence equations- based on the proceeded discussion and on data availability. Even though data on initial per capita GDP for the prefectures are available on a yearly basis (up to 1994 when per capita value added is published instead) for the Greek prefectures, this is not the case with other conditioning variables we wanted to consider. Data on education, employment in different sectors, etc. are given only for census years at this level of regional detail. Either we consider these variables at the beginning of a sub-period or taking their changes within a time interval, only in cross-section analysis can be used. The same is true for flow variables, such as private investment, which are given for one or few only sub-periods. Hence, even panel data analysis could not be applied using all conditioning variables, or if applied, it could not offer much given the very limited number of sub-periods for which the data are available within the period examined.

### 3. Econometric Estimates

Most studies on convergence conducted using cross-section data, estimate the equation form in (2) as such or augmented with conditioning variables, using non-linear least squares methods (NLS) even though ordinary least squares (OLS) is feasible. The purpose of applying NLS is to take T under consideration in a way that allows valid comparisons of the estimated parameter results and values of  $\beta$  between periods of different length. However, the same possibility exists if one estimates using OLS the following version of the model, which we finally estimated:

$$\frac{1}{T} \log \left( \frac{y_{i,t_0+T}}{y_{i,t_0}} \right) = a - (1 - b_T) \log(y_{i,t_0}) + u_{i,t_0,t_0+T} \quad (3)$$

where, as we conclude from (2) and (3), we can find  $\beta$  from  $(1 - b_T) = (1 - e^{-\beta T})/T$ . As is most often the case in the literature for obvious reasons, (3) was initially estimated as such, with no additional variables. The independent variable here was the logarithm of the per capita gross domestic product (GDP) of each prefecture at the beginning of the period. The dependent variable was the logarithm of the ratio of the annual per capita GDPs (in the first and last year of the period), divided by the number of years in the period. We estimated (3) for two overlapping periods. The first was the decade 1981-1991 for which investment data were available, and the second was the whole twenty-one year period 1970-1991. Sources of data were the relevant publications of the Centre of Economic Research and Planning (KEPE) and the National Statistical Service of Greece (NSSG).

First of all, it should be made clear that in all periods examined, there has not been  $\sigma$ -convergence in the Greek economy at the level of prefecture. On the contrary, the dispersion of prefectural per capita GDPs has increased over the years. The standard deviation for example, of the distribution of per capita GDPs for the fifty two prefectures in the years 1970, 1981, and 1991, was 0.006, 0.045, and 0,187 respectively. The trend towards greater dispersion of per capita GDPs therefore is clear. Nevertheless, a different story appears when one examines  $\beta$ -convergence using (3).

Table 1 below presents the results of the two estimations of (3), each for each period. It includes the estimated values of the coefficient of the independent variable with their t-values in the brackets, the corresponding derived values of  $\beta$ , and the values of  $R^2$ .

**Table 1: Econometric Estimates of the Convergence Equation**

Period	$\beta$	Estimated coefficient of $\log (y_{ist0})$	$R^2$
1981-1991	0.027	-0.024 (-4.24)	0.25
1970-1991	0.028	-0.021 (-5.52)	0.38

The estimated positive values of  $\beta$ , derived from the negative in all cases coefficients of  $\log(y_{i,t_0})$ , ( $-[1-b_T] < 0$ ) with high statistical significance, demonstrate the existence of  $\beta$ -convergence. The relatively low values of  $R^2$  are not unusual in such cross-sectional, convergence equation estimates. Such values in general can reflect the significance of omitted factors, differences of which, between regional units at the initial time, may determine different steady states for the regional per capita GDPs (the high level of statistical significance we found in the estimates of the constant term -not shown here- in all regressions, seems to advocates for the fact that the constant term picks up to a large extent the impact of other factors on the variation of the dependent variable).

The existence of other than the initial per capita GDP variables, which are significant and improve the explanatory power of the model, indicates that convergence is not absolute. Table 2 presents best results and econometric estimates of augmented versions of the model used, where additional variables have been considered and added to the model. Not all chosen variables or their combinations add substantially or the same to the performance of the model. Variables added linearly to the original model and appearing also in the table below, include the percentage share of the total work force employed in the primary sector (EPS), the percentage share of the total population which has completed formal secondary education (SEG) and the unemployment rate (UNEM). All these variables were taken first, at the beginning of each period representing the initial states. Another variable is the level of total private investment for the whole period, but at the regional level examined here, this information was available only for the period 1981-1991. Values of investment are expressed in millions of Greek drachmas (1 euro = 340.71 Dr).

**Table 2: Econometric estimates of the augmented models.**

Period	Estimated variable coefficients and t values						R <sup>2</sup>
	$\beta$	Log( $y_{i,t_0}$ )	EPS	SEG	UNEM	INV	
1981-1991	0.039	-0.032 (-5.13)	-----	-----	-----	0.14 (2.82)	0.38
	0.039	-0.032 (-5.12)	-----	-0.010 (-1.05)	-----	0.12 (2.06)	0.39
	0.040	-0.034 (5.27)	-0.013 (-1.33)	-----	-0.13 (-1.21)	0.11 (1.89)	0.41
1970-1991	0.021	-0.017 (-4.02)	-----	-0.045 (-1.46)	-----	NA	0.40

There are several other variables chosen and considered, but they were of very low statistical significance and added nothing to the model's explanatory power. An "index of land morphology" for example, based on the proportion of the area in every prefecture characterized as mountainous according to the National Statistical Service of Greece criteria (800m and above), was always found to be insignificant, showing the fact that land morphology characteristics do not necessarily impose a constraint on the rate of growth. The size of population was also of very low significance. Other combinations of the same variables as in the table above, had even lower explanatory power or did not add any, to the original model.

With regards to the additional variables used, it can be easily understood why they are so often considered in studies of growth. However, it is hard to capture well enough the influential factors they represent (eg. rate of secondary education graduates or other indices, to measure quality of human capital), and to grasp and describe particularities of each region. They ignore also other non-tangible factors affecting growth rates. This difficulty accounts partly for the relatively low R's in such cross sectional studies. This may be one of the reasons also, that most of the empirical studies focusing on the detection, classification and measurement of  $\beta$ -convergence using the theoretical framework of Barro and Sala-i-Martin, are still based primarily on estimates of the original non-augmented model. This, despite the bias possibly introduced due to omitted factors, especially if real per capita GDPs are not accurate.

Additional factors are often examined separately or later, in order to detect the “sources of convergence” (see for example A. de la Fuente, 2002 and Sala-i-Martin, 1996).

A problem encountered in some cases of the augmented versions of the model was multicollinearity, and this is one of the reasons as well, that some other combinations of variables and the relevant estimates are not included in Table 2. Multicollinearity was not strong however in those versions of the augmented model presented here, as a series of auxiliary regressions between independent variables showed. A detected major source of multicollinearity for the whole period 1971-1991 happened to be the relationship between the variables SEG and EPS and it was realized that it was better to rid of EPS.

The first result to be noted here is that the significance and signs of the estimated coefficients of the logarithmic per capita GDPs in all periods, together with the sign and value of  $\beta$ 's, lead again to the solid conclusion that there has been  $\beta$ -convergence of the economies of Greek prefectures.

In general, the significance of EPS, SEG, and UNEM, is small. Before we turn to discussing the signs of the estimated coefficients of these variables, we should mention that in general, they added little to the explanatory power of the original non-augmented model. The same is true with regards to the best performing versions of the augmented model themselves, presented in Table 2. For example, the explanatory contribution of SEG or EPS (or even both), to estimated models for the period 1981-91 is both, small and very similar to each other.

A different story appears when one examines the significance and explanatory power of the investment variable (INV). The significance of INV with regards to the original non-augmented and tried augmented models without INV, was far from negligible. In all combinations of variables tried, INV maintained a high level of statistical significance and it was the variable to which every increase in  $R^2$  and adjusted  $R^2$  is attributed almost entirely. The introduction of SEG and EPS in the model for 1981-1991 raised only slightly the value of  $R^2$  and the values of adjusted  $R^2$  (in general,  $R^2$ 's

and the values of adjusted  $R^2$ s are very close in all our estimates). Evidence shows therefore, that almost the entire increase in the explanatory power of the model for 1981-91, can be attributed to private investment. It is highly probable that if data on private investment were available for the whole period 1970-91, the results would have led to similar conclusions on the significance of private investment.

The negative signs of two of the conditioning variables we used and in general, different signs from study to study are very common in the literature. This can be seen also in Levin and Renelt (1992) where a long list of conditioning variables used in different studies and the estimated signs are provided. Negative signs have in our view their explanation. Poorer regions for example, tend to have a higher share of their population with less formal education. Hence, since poorer regions with lower per capita GDP tend to grow at higher rates and the coefficient of the initial per capita GDP is negative, the coefficient of SEG tends to be negative too. On the other hand, a large amount and share of population with completed formal education at the initial state means better human capital available. As a result the coefficient of SEG tends to be positive. Similarly, factors affecting the signs of other variables in opposite ways can be thought of, although not necessarily the same factors as the ones mentioned. The way that opposing factors determine the sign of these variables, depends largely on the choice of the initial state.

In recent studies on the sources of convergence (e.g. A. de la Fuente, 2002) utilizing the Barro and Sala-i-Martin approach and the augmented version of their model, instead of the exogenous variables taken at the initial state, what is used is the change in the values of these variables over a considered period. This is justified on the grounds that these changes over the period affect the average rate of growth for this period and they can be treated as exogenous, in the same way that other exogenous factors such as private investments are treated. This raises of course the issue of causality for the new exogenous variables since their changes may be the result of the growth process. In this case however, the problem affects the alternative of taking the conditioning variable values at the initial state. In particular the issue of the arbitrariness of the initial state choice is raised again.

We have estimated too the augmented models, using as additional -to the initial per capital GDP- exogenous variables, the changes of the previously considered additional exogenous variables at their initial state. As before, several combinations of chosen variables were tried and we present our best estimates in Table 3. The results clarify the role of the variables involved in the process of growth and convergence, including matters raised by the estimated signs and discussed above, since variables do have the expected signs now. As DEPS and DSEG we define the changes in the share of workforce employed in the primary sector (EPS) and the population share with completed formal secondary education (SEG) respectively, between the initial and last year of each period. DPOP is the percentage change of the population size, a variable that was found to be somewhat more significant in these augmented versions of the model and included in the estimates of Table 3. The variable INV is defined as above.

We also performed Granger causality tests for DPES, DSEG, and DPOP, in order to deal with the causality issue even though the direction of causality is usually taken for granted in the similar studies we are aware of. Results of estimates and the relevant F-tests showed that the direction of causality is from these variables to the average rate of growth (in a couple of instances, in alternative models examined for the two periods, the causality was bidirectional) and therefore we proceeded accordingly.

**Table 3: Econometric estimates of the augmented models**

Period	Estimated variable coefficients and t-values						R <sup>2</sup>
	$\beta$	Log( $y_{i,t_0}$ )	DEPS	DSEG	DPOP	DINV	
1981-1991	0.045	-0.036 (-5.40)	-----	0.207 (1.47)	-----	0.14 (2.68)	0.41
	0.037	-0.031 (-5.33)	-----	-----	-----	0.13 (2.66)	0.38
1970-1991	0.030	-0.022 (-4.96)	-0.026 (-2.01)	-----	0.016 (1.94)	NA	0.46
	0.033	-0.024 (-5.10)	-0.206 (-1.51)	0.064 (1.24)	0.17 (2.07)	NA	0.48

The results provide stronger evidence for conditional  $\beta$ -convergence. In contrast to the augmented models of Table 2, there is a clearer increase in the explanatory power over the non-augmented model of Table 1, as measured by  $R^2$ 's and adjusted  $R^2$ 's.

The negative sign of DEPS confirms the expected, that larger reductions in the share of the workforce employed in agriculture are associated with higher average rates of growth (in the observations reductions - which occurred in all prefectures in all periods - have negative signs). This variable was significant at 5% level of significance in one case only, but there was some significance, which added to the explanatory power of the model, in the other cases presented here as well. Some statistical significance of low levels for DSEG improved somewhat the performance of the model. A larger share of population with completed secondary education is contributory to greater average rates of growth as the estimated sign of DSEG shows. Population increases are also associated with greater average rates of growth.

Even though the inclusion of the variables appearing in Table 3 increases the values of  $R^2$ 's and adjusted  $R^2$ 's of the original non-augmented models of Table 1, it is the inclusion of private investment for the period 1981-1991 that leads to the greater increase of explanatory power. Given the fact that INV is consistently of relatively high statistical significance and increases substantially the performance of all models in which it is included, we can expect that if data on this variable was available the result would have been a better performance of the models in the other periods as well.

The significance of private investment is an issue that has been raised again in empirical studies. Levine and Renelt (1991, 1992) in particular, having found similar results, not only they argued that it is the initial per capita income and the amount of investment that exercise the greatest influence on the average rate of growth, but they proposed also a robustness test for the variables used. The only variables that passed this test were the initial per capita income and investment (even though the robustness of the investment was relatively smaller). Even though their test has been characterized as strict (A. de la Fuente, 1997) their quantitative analysis was detailed and well founded. It has been argued also that to some extent, the level of investment



as a variable does capture the impact of the non-tangible factors that affect growth such as networks, institutions, etc.

#### **4. An Artificial Neural Network Investigation of Convergence**

One of the criticisms cited with regards to use of convergence equations is the fact that their form is based on a log-linear approximation of the necessary conditions of an optimization problem. The log-linearization and, more important, the assumption of a certain form of the production (and utility) function, play a crucial role in deriving the form of the equation we used so far and hence the empirical results. Under different conditions, the form of convergence equations may differ as well and it may be also highly non-linear. To free the form of convergence equations from such restrictions we propose the use of artificial neural networks in their estimation.

A Back-propagation Neural Network algorithm (BPN), of the feed-forward form of Artificial Neural Networks (ANN), was used. The BPN connects the independent and dependent variables in such a way, that the implied functional form can approximate accurately any true relationship between the dependent and independent variables, including those that involve strong non-linearities. Unconstrained models and other augmented ("constrained") ones with additional variables, were estimated. Problems in the process of econometric estimation some of which we referred to, also advocate for the use of ANNs. Even though ANNs have found many applications in social sciences including economics and finance, to our knowledge there has been no extension of their applications to the empirics of convergence.

##### *A Synoptic Formal Description of BPN*

Neural networks represent parallel computing processes, which in various ways simulate biological intelligence. The simulation of course is far from a complete modeling of the brain's functioning complexity but it takes under consideration part of the existing knowledge on this functioning.

Every input (independent variable here) is represented by a neuron (input neuron) and all input neurons are thought of being placed in a layer, the input layer. Input neurons

are not connected between themselves but they are all connected to all neurons of another ("above") layer, which is called the middle ("hidden") layer.

Again, the neurons of the middle layer are not connected between themselves but each one of them is connected to the neurons of another ("above") middle layer, and so on. There is no predetermined number of middle layers or neurons in the middle layers. The final layer however is called the output layer and is constituted by the output neurons each one representing an output (dependent variable). There can be more than one dependent variable therefore. Each connection of a neuron  $i$  to neuron  $j$  of the next layer is given a weight  $w_{ij}$ . Neurons are also called nodes. In addition to the input, middle and output layer nodes, the structure of BPN usually contains in the input and middle layers a node, which is not connected to the nodes of the previous layer but it is connected to all nodes of the next layer. Such a node, often called a "bias", produces ("fires") constantly a value of one and transmits it to all nodes of the next layer, transformed by the weights of the corresponding connections. The BPN we employ in this study has one bias node in the input layer and another one in the middle layer.

The BPN is fed with the available statistical data (actual inputs and outputs) and is trained until it finds and adopts those values for the weights of connections, with which, if it is fed again with the same set of inputs it can predict accurately on its own the actual output. It is said then, that the network has learned and captured through these weights the true relationship between inputs and outputs.

We present here in a general formal way, the operation of the BPN. We assume the existence of one middle ("hidden") layer and hence this BPN includes in total three layers. There have been several different descriptions of the BPN in the literature, some of them written especially for economists (Kastens, Featherstone, and Biere, 1995). The description adopted here is based on Freeman and Skapura (1992) where a more detailed explanation of the BPN's operation along the following lines is presented.

Assume that we have P vector-pairs  $(x_1, y_1), (x_2, y_2), \dots, (x_P, y_P)$  representing an unknown functional relationship  $y = g(x)$  where  $x \in \mathbb{R}^N$  and  $y \in \mathbb{R}^M$ . The purpose is to train the 3-layer now BPN of N input nodes (independent variables) and M output nodes (dependent variables), using the P vectors, to "learn" by adjusting its weights the true relationship  $g(\cdot)$ . In the BPN's presentation here, y and x symbolize input and output variables and not the same variables as above, in the discussion of theoretical models and econometric analysis.

The steps that summarize the BPN's operation are the following:

1) An input vector  $x_r = (x_{r1}, x_{r2}, \dots, x_{rN})^t$  is applied to the input layer of N nodes, (where t simply denotes the number of times this vector is applied to the input layer since during training it is applied several times). Every input node sends the received signal to all nodes of the middle layer adjusted by each particular connection and therefore all middle ("hidden") layers receive adjusted signals-inputs by all input nodes. The net input to the  $j_{th}$  node of the hidden layer (h) is given therefore by:

$$net_{rj}^h = \sum_{i=1}^N w_{ji}^h x_{ri} + \theta_j^h \text{ where } w_{ji}^h \text{ is the weight of the connection between the } i_{th} \text{ input}$$

node to the  $j_{th}$  node of the hidden layer and  $\theta_j^h$  is a "bias term" which demonstrates the existence of an additional neuron in the input layer which constantly produces the value 1 and its connection weights are subject to updating like all other weights. Such a node has been found to improve the performance of the BPN.

2) This net input "activates" the  $j_{th}$  node of the hidden layer and produces an output according to the output function  $i_{rj} = f_j^h(net_{rj}^h)$ . This process happens in all hidden layer nodes. Output functions can take one from several forms, but we usually adopt the sigmoidal one, taking into consideration its monotonicity, its being output-limiting and its differentiability.

3) Each hidden layer node sends to all output nodes its signal, which is adjusted in the process by the connection weights. In particular, each node of the output layer

receives as a signal the total input:  $net_{rk}^o = \sum_{j=1}^L w_{kj}^o i_{rj} + \theta_k^o$ . Here  $net_{rk}^o$  is the net input

to the  $k_{th}$  node of the output layer, coming from all nodes of the hidden layer  $j =$

$1, 2, \dots, L$ ,  $w_{kj}^o$  is the weight of the connection between the  $j_{th}$  node of the middle layer

and the  $k_{th}$  output node. A node producing constantly the value 1 is also included in the middle layer too.

4) All output nodes process their received input and produce their output. The  $k_{th}$  output node for example produces the output  $o_{rk} = f_k^o(net_{rk}^o)$  where  $f_k^o(net_{rk}^o)$  is the output function of the  $k_{th}$  output node.

5) The error between the output (estimated value of dependent variable) of the output node and the real one is given by  $\delta_{rk} = (y_{rk} - o_{rk})$  where  $y_{rk}$  is the real output (value of dependent variable) of the output node (dependent variable)  $k$ , which corresponds to the input pattern  $r$ . It is this error that the BPN “sees” and tries to minimize by updating the connection weights. In fact for all given input and output patterns, the BPN trains and updates weights minimizing half the sum of squared errors

(Generalized Delta Rule -GDR):  $E_p = \frac{1}{2} \sum_{k=1}^M \delta_{rk}^2$  (4). We need to define the following

quantity:  $\delta_{rk}^o = (y_{rk} - o_{rk}) f_k^{\prime o}(net_{rk}^o)$  for the output nodes, where the prime indicates the derivative of the transformation function with respect to  $net_{rk}^o$ , and similarly

$\delta_{rj}^h = f_j^{\prime h}(net_{rj}^h) \sum_k \delta_{rk}^o w_{kj}^o$  for the hidden layer nodes.

6) The weights of connections between the middle and output layer nodes are updated according to:  $w_{kj}^o(t+1) = w_{kj}^o(t) + \eta \delta_{rk}^o i_{rj}$  where  $\eta$  is a predetermined “learning rate” parameter and  $t$  denotes the time we implement the input and output pattern  $r$  until the BPN trains properly. As mentioned, the BPN does not finish training with one input and output pattern, before it moves to the next one. It goes each time through all data again and again (“runs”), adjusting the weights. The described procedure takes place for all input and output patterns used in the training process.

7) The weights of connections between the input and middle layer nodes are updated as:  $w_{ji}^h(t+1) = w_{ji}^h(t) + \eta \delta_{rj}^h x_i$ . The order of the updates on an individual layer is not important. The training stops when the BPN “learns” the relationship between inputs and outputs and, given the final values of the updated weights, application of all input patterns will produce from the output nodes the true output patterns. The error measurement minimized always during training for the above purpose is always the one given by (4).

The transformation functions for all middle layer nodes are very often of the sigmoidal form as in most BPN applications for the well-known properties of this function. Hence,  $f_j^h(\text{net}_{rj}^h) = (1 + e^{-\text{net}_{rj}^h})^{-1}$ . This is the functional form we adopted also in our study. Since sigmoidal transformations of weighted sum of inputs take place in the middle layer nodes, we adopted as usual, a linear transformation function for the output nodes. This means that an output node simply produces (fires) the sum of all signals sent to this node by all middle layer nodes including the bias, weighted by the respective connection weights.

The learning rate parameter is positive and usually less than or equal to 1. It is related to the minimization of (4) through a steepest decent algorithm. A larger value of the parameter will result in a speedier training process but we run the danger of obtaining a local rather than a global minimum of (4) when the BPN converges, and further training does not improve the weights and the accuracy of the BPN predictions of the given output patterns from the input ones. A relatively small parameter will result in a larger number of runs (iterations)  $t$  and training time, but it increases the chances of reaching BPN convergence and final weight values at a lower or global minimum of (4). After a point however, reduction of the parameter may be unable to provide better results and even unable to achieve convergence of the training process at all. In Freeman and Skapura (1992) the derivation of the formulas for the updating of the weights can be found as well.

The way the trained BPN operates in order to make forecasts, is the similar to every run of the training process, except of course that no updating of weights and no repetition of runs takes place (this also renders the concepts of learning rate and error tolerance, irrelevant). If new input data are fed into a BPN after its training, the output values produced for each fed observation of input values are derived based on the structure and determined weights of the network.

Input values fed to their respective input neurons are sent to the next layer neuron's modified by the finalized weights of each connection. The bias node of the first layer behaves similarly sending the value 1 modified by the weights. Every next layer's neuron processes the received sum of all modified inputs, using the neurons' output

(transformation) function and produces its own output (signal). Subsequently each neuron of this layer sends its own signal to all neurons of the next layer modified by the respective connection weights. The bias node of this layer behaves as the bias node of the previous layer. The next layer nodes receive the sum of modified signals of the previous layer's nodes including the bias node, and the process goes on until we reach the nodes of the output layer and derive from each one of them a value for their respective outputs (dependent variables here).

The process above, through which a trained BPN makes a forecast using fed input values, can be described easily by a formula (C.M. Bishop 1996). This formula describes in fact, the relationship between inputs (independent variables) and outputs (dependent variables). The structure of the BPN defines the specific functional form of this mathematical relationship, which has the capability as mentioned, to accurately approximate any true relationship including those with strong non-linearities.

Let us assume in particular, one middle layer (hence, three layer BPN as in our case),  $d$  independent variables,  $k$  dependent variables,  $M$  middle layer nodes (without considering the middle layer bias node) and the same sigmoidal transformation function  $f(.)$  for all nodes in the middle layer (and the simple linear transformation we mentioned above for the output nodes). Then, with a slight modification in notation for the sake of clarification, we have reduced the general relationship implied by the BPN and presented in the literature, to:

$$y_k = w_{k0}^{(2)} + \sum_{j=1}^M w_{kj}^{(2)} f(w_{j0}^{(1)} + \sum_{i=1}^d w_{ji}^{(1)} x_i) \quad (5)$$

where  $y_k$  is the output produced by the  $k_{th}$  output node (dependent variable),  $w_{kj}^{(2)}$  is the connection weight between the  $j_{th}$  node of the second layer and the  $k_{th}$  node of the next layer (in our case where the only dependent variable is the average rate of growth,  $k=1$ ), and  $w_{ji}^{(1)}$  is the connection weight between the  $i_{th}$  node of the input layer and the  $j_{th}$  node of the second layer. The zero values of  $j$  refer to the two bias nodes, one in the input and one in the middle layer. The relationship in (5) can be written (with a sigmoidal  $f(.)$ ) as:

$$y_k = w_{k0}^{(2)} + w_{k1}^{(2)} \left( \frac{1}{1 + e^{\gamma_1}} \right) + w_{k2}^{(2)} \left( \frac{1}{1 + e^{\gamma_2}} \right) + \dots + w_{kM}^{(2)} \left( \frac{1}{1 + e^{\gamma_M}} \right) \quad (5a)$$

where,  $\gamma_i = -(w_{i0}^{(1)} + w_{i1}^{(1)} x_1 + w_{i2}^{(1)} x_2 + \dots + w_{id}^{(1)} x_d)$

Determining the optimum values of connection weights is equivalent to estimating the coefficients of the relationships between inputs and outputs represented by (5). From (5) and (5a) we derive the marginal impact of a change in the input variable  $x_i$  on  $y_k$  as:

$$\frac{\partial y_k}{\partial x_i} = w_{k1}^{(2)} \frac{e^{\gamma_1}}{(1 + e^{\gamma_1})^2} w_{1i}^{(1)} + w_{k2}^{(2)} \frac{e^{\gamma_2}}{(1 + e^{\gamma_2})^2} w_{2i}^{(1)} + \dots + w_{kM}^{(2)} \frac{e^{\gamma_M}}{(1 + e^{\gamma_M})^2} w_{Mi}^{(1)} \quad (6)$$

where,  $\gamma_i$  is defined as above.

### Empirical Analysis Using the BPN

We have applied a BPN as the one described above with the structure we mentioned (one middle layer, one bias node in the input layer and one in the middle layer) in the estimation of convergence equations and the testing of predictability of these equations. The number of input nodes is determined of course by the number of independent variables in every equation. There is no clear cut rule as to the optimum number of middle layer nodes. There are rules of thumb, which we utilized, together with our own searching to see when the network performs best, and chose a number of middle layer nodes. There are other more formal methodologies to optimize the structure of the network, such as using a genetic algorithm during training. Given the number of variables involved in our study and our own searching there were no reasons to resort to such procedures or expect a significant, if any, improvement in the network's performance. The number of middle layer nodes in our BPN was 11 including the bias node of this layer (M=10).

We have trained the BPN using the same data and variables as in the case of regression analysis. The training took place for the overall period 1970-1991 and for the period 1981-1991 for which data on private investment were available at the level of prefecture. Results showed that in all periods, the BPN substantially outperformed regression analysis, even when in this comparison stricter criteria are applied on the

BPN performance. In the network we trained we adopted a learning rate equal to 1, as is the usual default setting in ANN software. However, we experimented conducting the training with other values smaller than one. We tried also several values, changing during training, based on the performance of the network during this training. It was clear that other than one values of the learning rate did not help to improve the performance of the network for the given data set.

We adopted also a 0.10 tolerance rate. That means that during the training process and the relevant adjustments of weights described above, the network was considering an output prediction as accurate if it was less than equal to 10% of the range of real output values of the training data set. Forcing the network to learn and predict very accurately the outputs of the training data sets, if possible, leads to "loss of generality". This means reduced ability to predict and estimate with sufficient accuracy output values corresponding to input values not belonging. In other words, learning to produce from inputs of the specific training data set the relevant outputs, leads to less accurate approximation of the true relationship between inputs and outputs. The tolerance rate adopted is a very usual one in BPN applications but before its adoption we tried alternative values as well with no benefit to the network's performance.

In the training process of BPN we utilized 90% of the data. Once training stopped and the weights of the BPN were determined, we utilized the rest 10% of inputs and outputs for testing the BPN's predictability. This means that these 10% of input observations were fed to the BPN and, given the determined values of weights, corresponding output values were produced. Then, the predicted outputs, were compared to the real output observations. Errors therefore were analyzed to examine the predictability of the BPN with data which were not used for its training (unlike of course the lenient econometric evaluation of performance, based on  $R^2$  s, which refers to data used in the estimation of regression coefficients anyway). It was this way that the performance of the network was tested. The ability of the network to predict accurately the annual average rate of growth, which is our dependent variable, based on unused in training values of the independent variables, indicates that it has sufficiently captured the underlying relationship.



We have trained the network of every unconstrained and augmented model, three times. Each time a different number of data representing 10% of the total facts was kept outside training and was used for testing. Thus, in three testing procedures of each model, 30% of the available facts was used in total. As a result, the trained network was asked to make a sufficient number of predictions based on input data it didn't know from training. The testing facts were being selected randomly after reshuffling the data.

We present below, statistical information on the testing performance of the BPNs applied on each model, in order to evaluate their ability to predict accurately. Table 4 in particular, examines the performance of the BPNs in the cases of the unconstrained model (for the two periods) where the only independent variable is the per capita income at the initial state. There are 5 observations, which were used for testing after each training session, yielding in total 15 testing results for the three training sessions of each unconstrained model.

Table 4 provides information on the average absolute percentage of prediction error, of the BPNs, over all 15 testing data for each period. It provides also the maximum and minimum absolute percentage prediction errors. The last three rows of Table 4 provide the same information, but as a percentage of the entire data range. It can be argued that when outputs within and between both, training and testing data sets are close to each other, it is only natural that a BPN after training will have such weight values that will result to predictions close to the real outputs during testing. This implies small percentage errors. Even though it can be very questionable to what extent this is happening in a specific data set, we have considered errors as percentages of the data range in order to have a stricter evaluation of the network's performance. Partly because of expressing the dependent variable in logarithms, their differences from observation to observation are relatively small. Of course, it is not always the case that errors as percentages of the data range are greater than percentage errors.

**Table 4. Information on prediction errors and the performance of BPN**

	<b><u>1970-1991</u></b>	<b><u>1981-1991</u></b>
--	-------------------------	-------------------------

Average absolute percentage error	<b>0.025</b>	<b>0.022</b>
Maximum absolute percentage error	0.061	0.098
Minimum absolute percentage error	0.004	0.002
(Average absolute error)/(range of dependent variable values)	<b>0.121</b>	<b>0.758</b>
(Maximum absolute error)/(range of dependent variable values)	0.286	0.316
(Minimum absolute error)/(range of dependent variable values)	0.019	0.008

---

The above table testifies to the very satisfactory level of the BPN's performance. The average percentage error in absolute values is only 2.5% for the 1970-1991 period and 2.2% for the 1981-1991 period. The fact that all predictions were quite accurate is also demonstrated by the fact that the range of the above, prediction percentage errors, is between 0.4% and 6.1% only. Even if the stricter criterion of errors as percentages of the whole range of dependent variable values is used, the performance remains strong. The average error is just 12.1% of the range for 1970-1991 and 7.58% of the range for 1981-1991 (even though there is no clear-cut rule, a rule of thumb usually adopted by relevant software to characterize a prediction accurate is to be less than 40% of the range). The table also shows the maximum and minimum values of errors as percentages of the range. Naturally, their distance is greater than the previous distance between the maximum and minimum percentage errors.

This performance becomes even more impressive if one considers the fact that refers to predictions the corresponding values of which (values of independent and dependent variables) were not used for training and estimation of weights. Had we constructed a table with information on prediction errors corresponding to data used for the BPN's training and estimation of weights, (as econometricians do when  $R^2$ 's are calculated with data used for the estimation of coefficients), the performance would appear more impressive as expected.

In fact, we are not aware of any published empirical study that attained with the traditional cross-section or time-series analyses similar performance, especially in the

estimation of unconstrained models. A very high value of  $R^2$  in econometric estimation would have been derived if prediction errors were of the same magnitude as those in the BPN's testing, let alone the magnitude of smaller prediction errors in the estimation of weights. Sometimes values of  $R^2$  for ANNs are presented too, based on the two definitions of this measure of fitness of data in regression analysis. This may seem appealing, especially for comparison purposes, between ANNs or between certain ANNs (BPN, etc) and other methodologies (regression analysis, etc). However this is implausible since the definition of  $R^2$  breaks down as we move away from traditional regression. The two definitions of  $R^2$  or adjusted- $R^2$  lead to different values and even their range is not confined between 0 and 1 any more.

The performance of the network implies that its structure and determined weights have captured at a satisfactory level the true relationship between the average rate of growth and the initial per capita GDP. It is also interesting to note that through the relationship implied by the network, the average rate of growth could be determined to a very large extent by the per capita GDP at the initial state, as implied by the derived convergence equation of the original neoclassical model.

Table 5 shows the performance of the network when augmented models for the period 1970-1991 were estimated. As in the case of econometric analysis, conditional variables were the level of education, and employment share of the primary sector, as defined above. The variables in brackets were used in addition of course to the logarithmic value of the initial per capita GDP. They appear above the columns of the testing results corresponding to each model they were included, together or alone.

**Table 5. Information on prediction errors and the performance of BPN using augmented models for 1970-1991**

	(EDU)	(EDU, EPS)	(EPS)
Average absolute percentage error	<b>0.029</b>	<b>0.047</b>	<b>0.029</b>
Maximum absolute percentage error	0.088	0.121	0.043

Minimum absolute percentage error	0.006	0.017	0.012
(Average absolute error)/(range of dependent variable values)	<b>0.147</b>	<b>0.233</b>	<b>0.139</b>
(Maximum absolute error)/(range of dependent variable values)	0.408	0.564	0.20
(Minimum absolute error)/(range of dependent variable values)	0.031	0.086	0.058

---

The network performed again very well with the conditioning variables. The average percentage error in absolute value was only 2.9%, when either the education or the employment share of the primary sector was used. This is still very small of course, but it doesn't indicate any better performance than the unconstrained model for the same period. In addition when both variables are used the average percentage error in absolute values of the three testing trials becomes 4.7%. Again a very small error, but the lack of any improvement over the unconstrained model is even more apparent.

The higher absolute percentage error when both conditional variables are used shows that combining them, does more to "confuse" rather than help the network, as it tries to adjust and determine its weights. Many of the statistical properties of ANNs, (such as the extend to which the use of independent variables statistically related more or less between themselves, creates problems in determining the optimum weights) have not been fully investigated. However, given the complex and changing relationships that conditioning variables at initial states can have with the average rate of growth, which often leads to conflicting conclusions, and the way that the algorithm of the BPN operates, the above phenomenon can occur.

The average error as a percentage of the whole range of the dependent variable is slightly increased for the period, when one of the two conditioning variables is used (14.7% and 13.9% compared to 12.1% in the first column of Table 4). It is even higher at 23.3%, as the percentage error we discussed, when both conditioning variables are used. When education or both conditioning variables are used, it is the maximum values of the errors as percentages of the dependent variable range that increase significantly, something which does not happen when only employment

share in the primary sector is the conditional variable. This is due to very few observations only, behaving as outliers when education is included in the model. In general, we could argue that the BPN performed very well again, especially when only one of the two conditional variables was used. The addition of the two conditional variables for 1970-1991 however, did not improve the performance of the BPN in repeated testing of the augmented models.

Table 6, similarly constructed, shows the testing results of augmented models for the period 1981-1991 when the impact of private investment can be tested. Using education and the employment share of the primary sector together or not, led again to very good testing results but it did not offer any improvement in performance over the unconstrained model, as is the case for the whole period. In any case, we limit here our presentation of testing results to those that refer to augmented models including the investment variable.

**Table 6. Information on prediction errors and the performance of BPN using constrained models for 1981-1991**

	(INV)	(INV, EDU)
Average absolute percentage error	<b>0.045</b>	<b>0.045</b>
Maximum absolute percentage error	0.105	0.102
Minimum absolute percentage error	0.002	0.006
(Average absolute error)/ (range of dependent variable values)	<b>0.150</b>	<b>0.150</b>
(Maximum absolute error)/ (range of dependent variable values)	0.343	0.333
(Minimum absolute error)/ (range of dependent variable values)	0.006	0.022

The first column refers to the augmented model for the period 1981-1991 where the only conditional variable is the level of investment, while in the case of the second column in addition to the initial level of per capita GDP two conditional variables were included. These were the level of investment and education (using EPS did not improve the results). The very satisfactory performance of the BPN becomes apparent

again. The inclusion of the investment variable led to an average absolute percentage error in prediction, of only 4.5%. The average error as a percentage of the dependent variable error was 15%. Remarkably similar results were obtained as the second column shows when education was a conditional variable too.

The inclusion of the investment variable in the analysis for the period 1981-1991 improved the results over the performance of the other augmented models (which we do not show here) for the same period, when only EDU or EPS or both were the conditional variables. However, it did not offer any improvement over the unconstrained model of the period 1981-1991.

As we did with the regression analysis we have tested the performance of the BPN when the augmented models are conditioned by percentage changes in certain variables rather than their values at the initial state. Table 7, examines two such constrained models, one for 1970-1991 and one for 1981-1991, but as we did in the case of econometric analysis we have adopted as independent variable not the level of education at the initial states but their change over the whole periods. In particular, the first constrained model (1970-1991), contains as independent variable, in addition to the logarithm of the initial per capita income, the percentage change in the level of education. The second constrained model (1981-1991) includes, in addition to the logarithm of the initial per capita GDP, the level of investment for the period as before and the percentage change in the level of education.

**Table 7. Information on prediction errors and the performance of BPN using percentage changes in educational level as a conditional variable.**

	<b>1970-1991</b>	<b>1981-1991</b>
	(DEDU)	(DEDU, INV)
Average absolute percentage error	<b>0.020</b>	<b>0.038</b>
Maximum absolute percentage error	0.081	0.103

Minimum absolute percentage error (Average absolute error)/ (range of dependent variable values)	0.003 <b>0.093</b>	0.002 <b>0.128</b>
(Maximum absolute error)/ (range of dependent variable values)	0.358	0.337
(Minimum absolute error)/ (range of dependent variable values)	0.014	0.006

---

The results in the first column of Table 7, if compared to the first column of Table 4, show that the inclusion of the percentage change of education led to somewhat better testing results than those of the unconstrained model. The average absolute percentage error declined from 2.5% to 2% only, while the error as a percentage of the output value range declined from 12.1% to 9.3% (even though the value range of the two performance measures itself, has increased). The testing results were improved also, compared to those of the augmented model that includes education at the initial state. This can be seen comparing the first column of Table 7 to that of Table 5, and similar comments as above in the comparison with the unconstrained model, apply again.

When the percentage change of education is used in the constrained model of 1981-1991 together with the level of investment, there is clearly an improved performance in comparison to the augmented models for the same period where investment alone or together with education at the initial state are used. This conclusion is derived when the second column of Table 7 is compared to Table 6, and both measures of testing performance that we use, are examined. However, when the comparison is with the second column of Table 4, it becomes clear that the persistently very good testing performance of the BPN is not an improvement over the unconstrained model of 1981-1991.

### Estimated Relationships

After testing the performance of the BPN adopted, we trained it again using all data with the same learning rate and error tolerance. We estimated the connection weights

and therefore the parameters of (7) and (8). In order to avoid a lengthy presentation of estimates and, most important, given the very satisfactory and superior performance of the unconstrained model for 1970-1991, we present here this model as it stands after the estimation of connection weights. There are two input nodes (the logarithm of the initial per capita GDP and the input layer bias node), one output layer node (the average rate of growth as in the linear convergence equation), and the eleven middle layer nodes including the bias. This structure implies the existence of a (10x2) matrix of connection weights between the input and middle layer and a (11x1) vector of connection weights between the middle and output layer. There are therefore 31 weights estimated, entering as coefficients the estimated model. Resorting now back to the notation of the econometric model where  $y$  denotes per capita GDP, and based on the form of (5) and (5a), in our case this unconstrained estimated model becomes:

$$\begin{aligned} \frac{1}{T} \log \left( \frac{y_{i,t_0+T}}{y_{i,t_0}} \right) = & w_0^{(2)} + w_1^{(2)} \left( \frac{1}{1 + e^{-(w_{10}^{(1)} + w_{11}^{(1)} \log(y_{i,t_0}))}} \right) + w_2^{(2)} \left( \frac{1}{1 + e^{-(w_{20}^{(1)} + w_{21}^{(1)} \log(y_{i,t_0}))}} \right) \\ & + w_3^{(2)} \left( \frac{1}{1 + e^{-(w_{30}^{(1)} + w_{31}^{(1)} \log(y_{i,t_0}))}} \right) + \dots + w_{10}^{(2)} \left( \frac{1}{1 + e^{-(w_{10,0}^{(1)} + w_{10,1}^{(1)} \log(y_{i,t_0}))}} \right) \end{aligned} \quad (7)$$

From (7) we can obtain also the impact of a marginal change in the per capita GDP at the initial state, on the average rate of growth as:

$$\frac{\partial \left( \frac{1}{T} \log \left( \frac{y_{i,t_0+T}}{y_{i,t_0}} \right) \right)}{\partial (\log(y_{i,t_0}))} = \left( \frac{w_1^{(2)} e^{\gamma_1} w_{11}^{(1)}}{(1 + e^{\gamma_1})^2} \right) + \left( \frac{w_2^{(2)} e^{\gamma_2} w_{21}^{(1)}}{(1 + e^{\gamma_2})^2} \right) + \dots + \left( \frac{w_{10}^{(2)} e^{\gamma_{10}} w_{10,1}^{(1)}}{(1 + e^{\gamma_{10}})^2} \right) \quad (8)$$

where  $w_{\mu,1}^{(1)}$  ( $\mu = 1,2,\dots, 10$ ) is the connection weight between the initial per capita GDP (node 1 at the input layer) and node  $\mu$  of the input layer, and  $\gamma_\mu$  is defined as above (where  $i$  was used to denote nodes rather than  $\mu$ ). In our case:

$$\gamma_\mu = -(w_{\mu 0}^{(1)} + w_{\mu 1}^{(1)} \log(y_{i,t_0})).$$



The relationship given in (7) leads to rates of growth, which are not constant as in the case of log-linear convergence equations, but as (8) shows depend on the particular value of the initial per capita GDP. Hence, although the form of (7) refers here to cross-section analysis, it does provide information on the underlying processes, since a regional initial level of per capita GDP can be another region's initial per capita GDP at another time.

Inserting the estimated elements of the weight matrices into their corresponding places in (7) we have the estimated relationship (the two matrices are available upon request to the authors). Estimated weights confirm the negative relationship between the initial per capita GDP and the average rate of growth. For every initial per capita GDP used in the BPN training and weight estimation process, the estimated relationship of BPN yields average rates of growth which are consistently higher (lower) the lower (higher) is the initial per capita GDP. For few only sequential values of initial per capita GDPs, the predicted average rate of growth remains the same and then the clear negative relationship resumes again as the initial per capita GDPs change further. The negative relationship is confirmed also, if in addition to the used in training initial per capita GDPs, we use other values between them to estimate average rates of growth.

There is only one exception to the above results and concerns the very largest observed value of initial per capita GDP. When this value is inserted in the estimated BPN (7), it produces an estimated average rate of growth a bit larger than the prediction based on the immediate smaller rate of growth, violating the negativity of the relationship holding for the smaller values of initial per capita GDPs. This exception refers to the prefecture of Attica including the capital city Athens. A widely used explanation in the literature for a positive relationship between relatively higher initial per capita incomes and rates of growth is the existence of infrastructure and associated positive externalities in these initially richer areas.

If we rank the observed pairs of inputs and outputs we fed into the BPN for training according to the size of the initial per capita GDPs, it can be seen easily that lower

initial per capita GDP does not always imply greater average rates of growth, even though this is the case in general. This makes it even more interesting that the BPN, capable of capturing strong non-linearities and being so accurate, extrapolates after training a function of average rate of growth with respect to initial per capita GDP which is monotonically decreasing according to the above, and with one exception, greater initial per capita GDP implies always lower than equal average rate of growth for the range of input data used. Together with the BPN's very satisfactory performance in fitting the data and especially during testing, the estimated weights and the predicted outputs for different input values, indicate the existence of non-linearities in the studied relationship.

As is the case with all relationship estimation algorithms, it is expected that the BPNs performance is more accurate and reliable within the range of input values utilized. However, we fed into the network several values of per capita GDPs up to 20% above and 20% below the largest and lowest values used in training. The predicted values of average rates of growth do follow the results discussed above. As we reduce the values of initial per capita GDPs the predicted average rates of growth increase. However, as was the case with the highest observed and used value of per capita GDP, further increases lead to increasing predicted average rates of growth even though the increases are not relatively large. It seems therefore, that starting with the largest observed and used value of initial per capita GDP, the estimate of (7) establishes a new pattern where the negative relationship between initial per capita GDPs and the average rate of growth is reversed, for relatively high initial per capita GDPs.

## **5. Conclusions**

Our study has shown the usefulness and possibilities of Artificial Neural Networks in studying the existence and characteristics of regional economic convergence, using data on Greek prefectures. The Back-propagation ANN used captured extremely well the relationship between per capita GDPs at the chosen initial state and the average rate of growth during the examined period. The existing and used data fit very well the estimated by the BPN relationship. The BPN performed also extremely well in

repeated testing, in which unknown to the trained BPN data on initial per capita GDPs were used to predict the respective average rates of growth. In addition, the ANN clearly outperformed the econometric analysis conducted using the same variables over the same period. A result of the functional relationship implied by the BPN and its ability to accurately approximate any true relationship between the variables involved, regardless of strong non-linearities, is that the impact of the initial per capita GDP on the rate of growth (i.e. the derivative of the latter with respect to the former) is not constant as in traditionally used convergence equations but depends on the varying level of initial per capita GDP. This can provide also some information for the process of growth and convergence.

A major conclusion of our econometric analysis, is that in general poorer prefectures tend to grow faster than richer ones. This, despite the fact that there is not  $\sigma$ -convergence and the dispersion of per capita GDPs measured by the standard deviation of their distributions has increased for the periods examined, implying that the speed of converge does not suffice to close the gap between the per capita GDPs of poorer and richer prefectures. The values of the estimated coefficients  $\beta$  in our econometric analysis confirm not only the existence of  $\beta$ -convergence but also its slow speed. Many of estimated values of  $\beta$ , especially those between the values 0.02 and 0.03 are similar to those found in other empirical studies for other regions and countries. There has been some discussion in particular for the value of 0.02 which appears in several studies of regional convergence with regards to different provinces and countries.

The estimated relationship using the BPN is of a different form and the calculation of a parameter value, which indicates the speed of convergence, is a more complicated matter. It is also subject to the theoretical model used, from which one derives the convergence relationship that the BPN approximates. Even though the logic of the Barro and Sala-i-Martin approach was more relevant to our study, the BPN could be used as easily with other versions of the neoclassical approach such as the Mankiew, Romer and Weil approach, if data on additional variables needed were available. The estimated relationship confirms too, the generally negative relationship between the initial per capita GDP and the average rate of growth. It confirms also that this

relationship is weak in the sense that changes in the initial per capita GDP do not bring about dramatic changes in the average rate of growth and in some cases the changes are close to zero. It is interesting however, that with the BPNs ability to capture possible non-linearities and its strong performance, the neural network sees clearly and certifies a negative relationship in each prediction it makes using the actual data on initial per capita GDPs. And this, despite the fact, that in these data there are certain violations of the negativity of the relationship.

However, unlike the results provided by traditional econometric analysis, the above results hold up to a point. For the very richest prefecture at the initial state, the trained BPN yields an average rate of growth higher than the rate corresponding to the five observations with the immediate lower initial per capita GDPs. It appears that starting with a value around the one of the highest initial per capita GDP, the BPN yields a positive relationship between the initial per capita GDP and the average rate of growth. Taking under consideration the observed exceptions to the negativity of the relationship, it is nevertheless, the particular observation with the highest initial per capita GDP that “convinced” the BPN to provide a positive relationship at relatively high values of the independent variable.

As it is natural, the trained BPN provides better data fitting and predictability within the range of the independent variables used in training. However, predictions of the trained BPN using values of initial per capita GDP below and above the training range of this variable, lead to results that confirm the above patterns. The trained BPN therefore facilitates the conclusion that poorer prefectures tend to grow faster than richer ones, with this negative relationship between initial per capita GDP and average rates of growth turning into a positive once initial per capita GDP is taken at a relatively very high level.

In the econometric analysis a positive and significant relation between private investment and the average rate of growth was established, even though data availability at the level of prefecture allowed us to examine the role of investments for a period only. The significance of such findings in terms of policy implications are profound. Private investment seems to be not only a major force towards greater rates

of growth but the most effective tool if the desire is to raise the speed of convergence and reduce the observed increasing regional disparities in terms of per capita GDPs. Results show also that higher rates of growth are associated with reductions in the share of the workforce employed in the primary sector. Higher rates of growth are associated also with increases in the share of population with completed secondary education and the size of population, but its role was certainly less significant than that of private investment and initial per capita income. Other factors were found to add very little or nothing to the explanation of average annual rates of growth. There are of course factors, which are difficult to incorporate in the analysis either because they are non-tangible or due to lack of data at the regional level examined (infrastructure at the initial time public investment, etc), even though some of the non included factors may be strongly correlated with variables considered.

In the case of the more flexible and accurate BPN, it appears that in general, even the inclusion of investment does not offer any assistance in improving the BPN's performance. Unconstrained models do a very good enough job on their own. This is not so surprising considering econometric studies which estimate augmented models, and where investment is the only conditional variable passing robustness tests, and not so clearly. This is not to say that investment (or other factors as well) is not a major player in the growth and possible convergence process. It is simply possible that their role can be sufficiently captured by the relationship of the initial per capita income and the average rate of growth, as this is estimated using a BPN. It should be remembered also that the derivation of unconstrained convergence equations such as the ones estimated here, does take under consideration within the neoclassical theoretical framework, the formation of capital which is the product of investment. This is the case for other variables too, which in econometric analysis are used often as conditioning variables.

Given the problems arising in econometric analysis (biases, etc), the relatively poor performance of econometrically estimated models and the low  $R^2$ s, the results on the significance of conditioning variables should be treated with skepticism. This should be the case even if there were not the results of a more efficient and reliable tool of analysis as the trained BPN appears to be. Conditioning variables can be indeed

significant in cases where they represent or pick up the impact of those factors that influence rates of growth and differ substantially from region to region. Examples are different monetary and fiscal policies, different trade regimes, etc. In general, inability to capture the influence of policy and other external factors, seriously undermines the derivation of conclusions using empirical analysis on whether there is a tendency towards convergence or not, which is inherent to the economic system. One hopes that when examining regional economies subject to similar kind of interventions, policies, legal and institutional frameworks, the problem is substantially reduced.

Finally, the ability of the BPN to capture existing non-linearities in the relationship between initial per capita incomes and average rates of growth, allows the analysis of cross-sectional data to provide useful information on the dynamic process of growth and convergence. This was evident in the case of our data set, where together with the changing strength of the relationship we saw also its reversal.

## References

Arrow, K. J., 1962, "The Economic Implications of Learning by Doing", reprinted in: K. J. Arrow, 1985, *Collected Papers of Kenneth J. Arrow*, Vol. 5, 157-180, ("Production and Capital"), Harvard University Press, Cambridge, Massachusetts.

Abramovitz, M., 1979, "Rapid Growth Potential and its realization", reprinted in: M. Abramovitz, 1989, *Thinking About Growth and Other Essays on Economic Growth and Welfare*, 187-219, Cambridge University Press, Cambridge.

Abramovitz, M., 1986, "Catching Up, Forging Ahead and Falling Behind" reprinted in: *Thinking About Growth and Other Essays on Economic Growth and Welfare*, 220-244, Cambridge University Press, Cambridge.

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

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

Barro, R., 1997, "Determinants of Economic Growth", MIT Press, Cambridge.

Benhabib, J. and Spiegel, M. M., 1997, "Cross-Country Growth Regressions", Working Paper 97-20, CV Starr Center, New York University.

Bernard, A. B. and Durlauf, S. N., 1995, "Convergence in International Output", *Journal of Applied Econometrics*, 10(2), 97-108.

Bernard, A. B. and Durlauf, S. N., 1996, "Interpreting Tests of the Convergence Hypothesis", *Journal of Econometrics*, 71(1-2), 161-174.

Bianchi, M., 1997, "Testing for Convergence: Evidence from Non-Parametric Multimodality Tests", *Journal of Applied Econometrics*, 12(4), 393-409.

Bickenbach, F. and Bode, E., 2003, "Evaluating the Markov Property in Studies of Economic Convergence", *International Regional Science Review*, 26(3), 363-329.

Bishop, C. M., 1996, "Neural Networks for Pattern Recognition", Oxford University Press.

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

Caselli, F., Esquivel G., and Lefort, F., 1996, "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics", *Journal of Economic Growth*, 1(3), 363-389.

Cass, D., 1965, "Optimum Growth in an Aggregative Model of Capital Accumulation", *Review of Economic Studies* XXXII, 223-240.

- Chamley, C., 1981, "The Welfare Cost of Capital Income Taxation in a Growing Economy", *Journal of Political Economy*, 89, 468-496.
- de la Fuente, A., 2002, "On the sources of convergence: A close look at the Spanish regions", *European Economic Review*, 46 (2002) 569-599.
- de la Fuente, A., 1997, "The Empirics of Growth and Convergence: A Selective Review", *Journal of Economic Dynamics and Control*, 21, 23-73.
- Durlauf, S. N., 1989, "Output Persistence, Economic Structure, and the Choice of Stabilization Policy", *Brookings Papers on Economic Activity*, 2, 69-116.
- Durlauf, S. N. and Johnson, P. A., 1995, "Multiple Regimes and Cross Country Growth Behavior", *Journal of Applied Econometrics*, 10(4), 365-384.
- Durlauf, S. N., 1993, "Nonergodic Economic Theory", *Review of Economic Studies*, 60(2), 349-366.
- Durlauf, S. N., 1996, "A Theory of Persistent Income Inequality", *Journal of Economic Growth*, 1(1), 75-93.
- Durlauf, S. N. and Quah, D. T., 1998, "The New Empirics of Economic Growth", The University of Wisconsin, Madison and LSE.
- Duffy, J. and Papageorgiou, C., 1997, "The Specification of the Aggregate Production Function: A Cross Country Empirical Investigation", Working Paper, University of Pittsburg.
- Evans, P., 1996, "Using cross-country variances to evaluate growth theories", *Journal of Economic Dynamics and Control*, Vol. 20, No, 6-7, 1027-1049.
- Evans, P. 1998, "Using Panel Data to Evaluate Growth Theories" *International Economic Review*



- Evans, P. and Karras, G., 1996, "Convergence revisited", *Journal of Monetary Economics*, 37, 249-265.
- Franses, P. H., and Hobijn, B. 1995, "Convergence of Living Standards: An International Analysis", Technical Report 9534/A, Econometric Institute, Erasmus University Rotterdam.
- Freeman, J. A., and Skapura, D. M., (1991), "Neural Networks: Algorithms, Applications, and Programming Techniques", Addison-Wesley, Inc.
- Friedman, M., 1992, "Do Old Fallacies Ever Die?", *Journal of Economic Literature* XXX, 2129-2132.
- Grossman, G. and Helpman, E., 1991, *Innovation and Growth in the Global Economy*, MIT Press, Cambridge, Massachusetts.
- Islam, N. 1995, "Growth Empirics: A panel Data Approach", *Quarterly Journal of Economics*, 110(4), 1127-1170.
- Johansen, S., 1988, "Statistical Analysis of Cointegration Vectors" *Journal of Economic Dynamics and Control*, 12, 231-54.
- Johansen, S., 1991, "Estimation and Hypothesis Testing of Cointegrating Vectors in Gaussian Vector Autoregressive Models" *Econometrica*, 59, 1551-80.
- Jones, C. I., 1997, "On the Evolution of the World Income Distribution", *Journal of Economic Perspectives*, 11(3), 19-36.
- Jones, L. E., and Manuelli R., 1990, "A Convex Model of Equilibrium Growth: Theory and Implications", *Journal of Political Economy*, 98(5), 1008-1038.
- Kastens, T. L., Featherstone, A. M., and Biere, A. W., 1995, "A Neural Network Primer for Agricultural Economists", *Agricultural Finance Review*, 55, 54-73.

King, R. G., and Rebelo, S. T., 1989, "Transitional Dynamics and Economic Growth in the Neoclassical Model", NBER, Working Paper no 3185, Cambridge, Massachusetts.

Koopmans, T., 1965, "On the concept of optimal economic growth", in: *The econometric approach to development planning*, North-Holland, Amsterdam.

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

Levine, R., and Renelt, D., 1991, "Cross-Country Studies of Growth and Policy: Methodological, Conceptual, and Statistical Problems", World Bank, Working Paper no 608, Washington, DC.

Levine, R., and Renelt, D., 1991, "A Sensitivity Analysis of Cross Country Growth Regressions", *American Economic Review*, 82(4), 942-963.

Levin, A., and C.-F. Lin, 1992, "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties", Working Paper, University of California, San Diego.

Levin, A., and C.-F. Lin, 1993, "Unit root tests in panel data: New results", Working paper no. 93-56, University of California, San Diego.

Levine, R., and Renelt, D., 1992, "A Sensitivity Analysis of Cross-Country Growth Regressions", *American Economic Review* 82, 942-963.

Lucas, R., 1988, "On the Mechanics of Economic Development", *Journal of Monetary Economics*, 22, 3-42.

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

Nerlove, M., 1996, "Growth Rate Convergence, Fact or Artifact?" Working Paper, University of Maryland.

Nordhaus, W., 1969, "An Economic Theory of Technological Change", *American Economic Review*, 59, 18-28.

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

Quah, D. T., 1996a, "Empirics for economic growth and convergence", *European Economic Review*, 40, 1353-1375.

Quah, D. T., 1996b, "Convergence Empirics Across Economies With (some) Capital Mobility", *Journal of Economic Growth*, 1(1), 95-124.

Quah, D. T., 1994, "Exploiting Cross Section Variation for Unit Root Inference Dynamic Data", *Economics Letters*, 44(1), 9-19.

Quah, D. T., 1993a, "Galton's Fallacy and Tests of the Convergence Hypothesis", Discussion paper no. 820, CEPR, London.

Quah, D. T., 1993b, "Galton's Fallacy and Tests of the Convergence Hypothesis", *Scandinavian Journal of Economics*, 95, 427-443.

Quah, D. T., 1993c, "Empirical Cross-Section Dynamics in Economic Growth", *European Economic Review*, 37(2/3), 426-434.

Quah, D. T., 1992, "International Patterns of Growth: I. "Persistence in Cross-Country Disparities", Working Paper, LSE, London.

Ramsey, F. P., 1928, "A Mathematical Theory of Saving", *Economic Journal* 38, 543-559.

Rebelo, S., 1991, "Long-Run Policy Analysis and Long-Run Growth", *Journal of Political Economy*, 99(3), 500-521.

Romer, P., 1986, "Increasing Returns and Long-Run Growth", *Journal of Political Economy*, 94, 1002-1037.

Romer, P., 1990, "Endogenous Technical Change", *Journal of Political Economy*, 98, S71-S102.

Sala-i-Martin, X., 1996, "Regional cohesion: Evidence and theories of regional growth and convergence", *European Economic Review*, 40, 1325-1355.

Schultz, T., 1960, "Capital Formation by Education", *Journal of Political Economy*, 69, 571-583.

Schultz, T., 1961, "Investment in Human Capital", *American Economic Review*, 51, 1-17.

Shell, K., 1966, "Toward a Theory of Inventive Activity and Capital Accumulation" *American Economic Review*, Papers and Proceedings 56, 62-68.

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

Uzawa, H., 1965, "Optimum Technical Change in an Aggregative Model of Economic Growth", *International Economic Review* 6, 18-31.