STOCKHOLM SCHOOL OF ECONOMICS

Club Convergence in the European Union

Analyzing the impact of the financial crisis of 2008

Johannes Herkules Stefanoudakis
12/12/2011

ACKNOWLEDGEMENTS

I would like to express my gratitude to those who made this thesis possible. I am especially thankful to Christos Papahristodoulou, Professor at Mälardalens University and to my supervisor, Örjan Sjöberg, for their support and guidance until the completion of the thesis.

Lastly, I am grateful to all those who supported me by any means throughout the development of this project.

Johannes H. Stefanoudakis

ABSTRACT

JOHANNES HERKULES STEFANOUDAKIS: Club Convergence in the European Union: Analyzing the impact of the financial crisis of 2008.
(Under the direction of Örjan Sjöberg)

The aim of this paper is to empirically identify the impact of the financial crisis of 2008 on the mechanisms of within-club convergence of GDP per capita, for the 27 member countries of the European Union. To tackle this issue, the author first uses spatial filtering methods to identify club formation under a concept of bimodal distribution and then proposes a two-stage framework in which unconditional and conditional convergence are estimated before and after the crisis period. Regression analysis is performed with the use of ordinary least squares method. The results indicate that within club convergence rate slows down for the high income club after the hit of the financial crisis. However, the intriguing results of the low income club regression analysis make space for some very interesting inferences and set the basic foundation for future research.

Keywords: Club convergence hypothesis, European Union, Economic crisis, financial crisis, spatial dependences.

Table of Contents

List of Figures

Introduction

In 1951, six countries – Belgium, Germany, France, Italy, Luxembourg and the Netherlands – founded the European Coal and Steel Community. Through the years, various countries of the European continent that met specific requirements were approved to join this community which slowly expanded to include 27 member countries under the name of the European Union. The core definition of this union is that "…it is a unique economic and political partnership between 27 European countries"[1]. In order to be granted membership to the union, countries should fulfill the so called "Copenhagen criteria", established in 1993. These can be summarized in the following three points[2].

- Stable institutions that guarantee democracy, the rule of law, human rights and respect for and protection of minorities.
- A functioning market economy, as well as the ability to cope with the pressure of competition and the market forces at work inside the Union.
- The ability to assume the obligations of membership, in particular adherence to the objectives of political, economic and monetary union.

Collecting the economies of many countries with similar characteristics (as was assumed) under a single union could improve the economic position of the European Union. In addition, with the introduction of a common currency the target was, and still is, to promote growth and strengthen trade relations between countries and with the rest of the world, by minimizing the differences between them. The growth theories that were established long ago by Solow and later on by Romer, characterize this minimization of differences as convergence. There has been a vast amount of empirical studies aiming at exploring the evolution of such output differences among countries. The empirical results illustrate that instead of observing overall convergence of the economies within the European Union countries tend to cluster together, creating economic clubs. This being the case, the convergence hypothesis is valid for the countries belonging in the same club but not for the union as a whole.

[1] The information on the characteristics of the European Union was obtained from the union's official website: http://europa.eu/about-eu/basic-information/index_en.htm, last visited: 2011-11-29.
[2] http://ec.europa.eu/enlargement/the-policy/conditions-for-enlargement/index_en.htm, last visited: 2011-11-29.

The recent events of the collapse of the financial system in 2008 have caused the creation of an economic crisis in the years leading to today. Most of the countries' economies have plunged during 2009 experiencing negative effects in terms of innovations, education, unemployment, trade, consumption and consequently growth. Despite the evident recovery of countries in 2010, a great degree of economic instability has persisted and increased within the European Union. The main question that arises at this point is how club formation and convergence within the club has been affected by the economic crisis?

This research papers addresses this question and tries to understand the effects (if any) of economic shocks, on the convergence rate within the clubs formed. The contribution of this research paper to the existing literature on club convergence is two-fold. It first aims at extending the time interval of older studies to 2010 and then by taking into account the recent economic downturn from 2008 to today, capturing the impact of the crisis on the mechanisms of club convergence. This research paper will address the following questions: What determines club convergence? Do empirical data still illustrate club convergence in the aftermath of the financial crisis of 2008? If so, which clubs are formed and have they changed? What is the impact of an economic crisis on within club conditional convergence? Have any signs of economic slowdown in the convergence rates within the clubs appeared? How are the structural characteristics of countries impacted by the crisis?

In order to effectively address the above mentioned questions, this research paper has been structured as follows. The first section contains the methodology that was used to approach, understand and analyze the problem at hand. That is, underlying the best way to study the impact of the economic crisis on club formation and convergence within the clubs. Defining the scope of the research was highly important and thus the second section includes all the aspects that even though they might affect or even change the outcome of the analysis, were considered but not analyzed during the empirical testing.

The third section sets out the theoretical foundation of this research paper, the theory behind the mechanisms of growth and convergence. It starts by underlying the origins of growth theory and the concept of convergence within the neoclassical growth model. It provides a detailed description of the three types of convergence (unconditional, conditional and club convergence) and proceeds by highlighting the findings of previous empirical literature regarding club convergence. The theoretical section also emphasizes the importance of spatial modeling and presents the two main methods for modeling spatial dependence. The last part of this section discusses the impact of “shocks” on economic growth highlighting their nature and channels of impact.

The fourth section is maybe the most crucial step before the empirical testing. This is due to the fact that it contains the three models used for empirical analysis. It starts with a focus on the statistics and measures used for the definition of the clubs formed and spatial modeling. It provides a detailed description of how Moran’s I and Getis’ and Ord’s G statistics are used in the identification of clubs formed and country membership. It then continues with presenting the regression equation and the model of beta convergence testing under both the unconditional and conditional assumption. This fourth section concludes with an overview of the variables used during regression analysis as well as their expected behavior.

The fifth and last section before the conclusion presents all the findings of the empirical analysis and provides the reader with some inferences regarding those findings and the behavior of the variables used. To be more precise, it starts by presenting the findings of empirical testing under the unconditional framework for both the high and low income club countries, as well as making some in-depth analysis of those first findings. It then proceeds by presenting the regression results within the conditional convergence framework, again for the high and low income club. Containing a more detailed analysis, each club’s conditional convergence analysis is followed by a section of inferences and interpretations of the findings. Last but not least the findings of this research are summarized in the “conclusions” part.

Compared to previous studies of club convergence, the uniqueness of this research paper lies in three specific points. The first is that it examines club convergence at the national level whereas most of the empirical literature has been focusing on the regional level analysis. Secondly it clearly defines an appropriate measure for capturing the impact of spatial influences (measuring the value of δ). Last and most importantly it incorporates a period of economic turbulence in the time period of study allowing for the study of economic shocks. Even though the time interval that accounts for the crisis is relatively short (two years), we expect to find some differences in trends among the clusters formed. Being the most up to date study of club convergence in the European region and the only one to account for the years after the financial crisis, we expect that it will bring additional information in the club convergence debate.

Methodology

In order to be able to address the research questions presented above and succeed in conducting a thorough analysis of the club convergence hypothesis, the following course of action was designed and followed throughout the report. Needless to be said, the first step was to conduct an in depth research of the theoretical frameworks and existing empirical literature underlying the growth, convergence and club-convergence of economies.

Since empirical analysis is at the core of this paper, when defining the scope of the research, a decision had to be made on the boundaries of the data obtained. First and most importantly, defined also by the title of this research paper, the set of countries had to include only the member states of the European Union[3]. In addition, the time interval of study had to be long enough in order to address two potential issues. The first issue with short time series is that when economies are close to their steady state, deviations reflect mainly cyclical and structural shocks rather than convergence (Tselios, 2009). The second potential issue was that of incomplete data series. For most of the countries, before 1995 there are a lot of data gaps and for the low income countries there are data missing between 1995 and 1999. In order to conduct a thorough analysis, it is crucial to obtain values for all the variables, time intervals and countries. For these reasons, an interval of 15 years was chosen ranging from 1995 to 2010. These data series were obtained from EUROSTAT (European Commission's statistical database).

For the empirical analysis itself the first step is to define the clusters formed with the use of explanatory spatial data analysis tools. The local clustering technique presented by Getis and Ord (1992) will be used for that estimation and then analyzed together with Moran's I statistic to give an in depth insight into the patterns formed. With the use of normalized G-statistics the overall sample is split under the concept of bimodal income distribution, creating two clubs; a higher income club and a lower income club. More details regarding the statistics of this method are provided on the sub-section labeled "Modeling Spatial Dependence". Then, convergence is tested for the countries within each group (separately for each club) with the use of both the unconditional and conditional β-convergence model. A detailed description of the models is presented in sub-section "Beta Convergence Modeling".

[3] A list of the countries taken into account is presented in the Appendix 1.

In order to be able to capture the impact of the economic downturn on club-convergence and understand its mechanisms, a two-stage framework was designed. As illustrated in the figure below, with starting year 1995, the club convergence hypothesis is first examined for the years ending in 2008. When addressing club convergence the focus will be put on the factors that may cause convergence (initial conditions, structural & spatial characteristics), the degree of influence and the clusters formed.

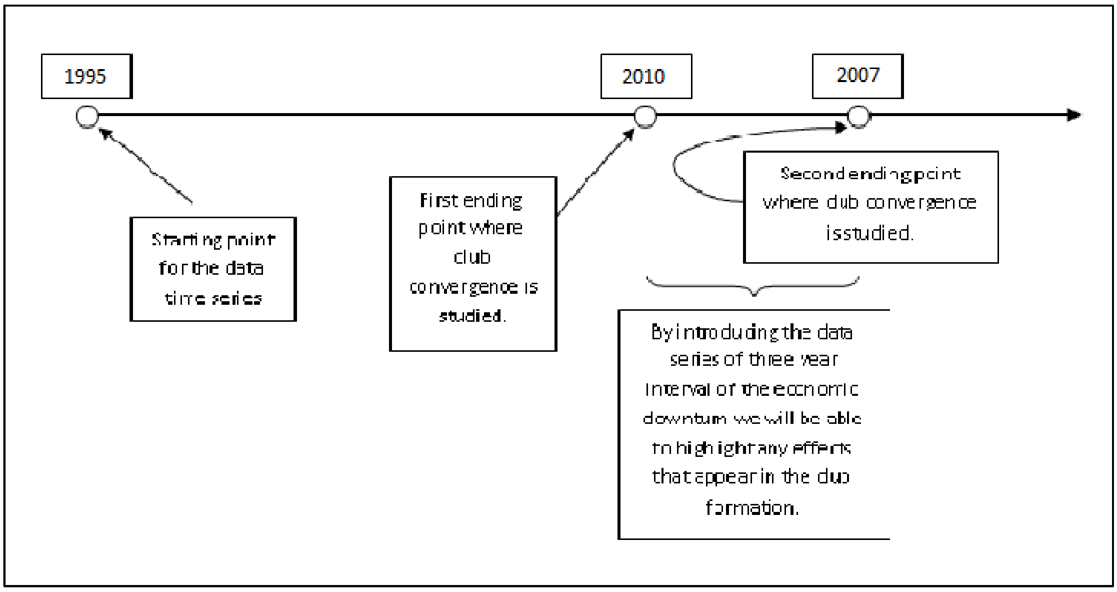

Figure 1: The timeline of the club convergence study
Source: Own illustration.

The next stage entails the exact same study as before but this time after extending the time series to 2010. Last but definitely not least inferences regarding the behavior of the clubs are made, after comparing the findings from two points in time (before and after the crisis). By adding the years of 2009 and 2010, differences attributed to the economic downturn can now be monitored and analyzed. In terms of findings however, within club convergence is not expected to change dramatically. Even though divergence may occur, it is a rather extreme outcome and what it is expected to be seen is a slowdown in the convergence rates of each club.

Scope of the Study

In defining the scope of the research paper, it is highly important to provide sufficient explanations as to why certain factors are not taken into consideration as well as underline the limitations of the empirical study. To begin with, it is not going to be evaluated whether or not a country is at its steady state level since such computations entail many difficulties and lies outside the scope of this report. However, a way to address this issue would be to analyze the correlation between initial levels of national income and subsequent growth rates (Fischer & Stirböck, 2006).

As mentioned before, starting year for the empirical study is 1995. This choice relates solely to the fact that data availability and accuracy before that year decreases significantly. Due to certain assumptions made in this report, convergence will only be tested at a national level of countries. Acknowledging the importance of sub-national variations as important determinants of growth and convergence, it is vital to mention that the best way to capture those variations is with the use of data at a regional level. However for reasons presented below, national data analysis is chosen instead. First and most importantly restrictions on data availability prohibit the use of regional data. Regional data extend only to 2008 in the EUROSTAT database, whereas the latest data on a national level extend to 2010. In order to be able to follow the two-stage analysis presented in the previous section, data from as late as 2010 must be available for testing, thus narrowing down our choice.

Even though examining convergence on a regional level can give further insight into club formation, it is important to point out some of the limitations of choosing a micro-level analysis. By moving away from the aggregate levels, more and more emphasis is put on intra-regional activities that determine long-run growth and less emphasis on the inter-regional activities which are equally important. Shifting the focus back to inter-regional activities, club convergence can still be studied by examining the nations themselves and taking into account the impact of spatial characteristics. Even though using data at the national level prohibits the analysis to capture the full effects of intra-regional variables (knowledge spillovers, agglomeration effects, etc.) (Fischer & Stirböck, 2006), it is still expected to provide a good insight into the dynamics of spatial dependencies.

Another advantage of shifting the focus back to the national level is that it minimizes significantly the modifiable areal unit problem (MAUP)[4]. It is also important to point out that regions (as defined by the EU) are rather formal than functional units (Fischer & Stirböck, 2006) and to add to that, the assumption of diminishing returns that drives the neoclassical convergence process and the assumption of a closed economy are particularly questionable for regional economies. Last but not least, growth will be considered as "generative" rather than "competitive", meaning that national growth is treated as the outcome of the growth rates comprising the economy (Richardson, 1978).

[4] The modifiable areal unit problem is composed by two parts, the scale problem and the zoning problem. The first one refers to the difficulties in choosing an appropriate spatial scale of analysis. The other one refers to how the spatial units are configured (Getis, 2005).

Theoretical Background

Convergence origins

At the basis of growth theory (based on the concept of the production factors[5]), there are two competitive theories. The first one, introduced by Solow (1956), is the neoclassical growth theory and is known as convergence optimism, supporting the idea of a convergence in income levels among countries. The later, introduced by Romer (1986), is the endogenous growth theory, also known as convergence pessimism. It supports the idea of a persisting and increasing income inequality in income levels. Being in a constant debate for years, both theories have received a fair amount of criticism for their findings. The neoclassical growth theory, on the one hand, fails to take into account the importance of entrepreneurship and institutions. In addition it cannot explain why or how technological progress occurs. Due to the later drawback, endogenous growth models were introduced, which in turn failed significantly in explaining conditional convergence reported in the empirical literature.

With empirical literature partially supporting the model, neoclassical growth theory has generated a great deal of research and discussion around the concept of income convergence among countries. Economists have tried to use existing theoretical models to predict and analyze the countries' growth behavior. The notion of convergence describes that under specific circumstances poor countries will tend to grow faster than rich ones, in order to narrow down the gap between them. Convergence can be seen from two perspectives (Tiiu, Kuusk, Schlitte, & Võrk, 2007):

- Convergence in terms of the steady state growth rate and,
- Convergence in terms of the level of income.

According to the former, convergence takes place in the sense that the steady state growth rate of countries will be the same in the long-run. In this research paper however, the focus will be on convergence in income levels among countries, implying that the steady state income levels of countries tend to be the same in the long-run.

[5] A different view studies economic growth on the basis of the institutional economic growth theory.

At the center of the neoclassical growth theory economists have underlined three competitive theories that try to explain the concept of convergence:

- The absolute (unconditional convergence hypothesis)
- The conditional convergence hypothesis
- The club convergence hypothesis

The absolute convergence hypothesis implies that countries or regional economies will converge with each other, regardless of their initial conditions. This model assumes convergence takes place even if no other explanatory variables are introduced to the model. Such a model would be a reasonable assumption for homogenous samples (USA states, OECD, European regions) (Tiiu, Kuusk, Schlitte, & Võrk, 2007). However ignoring other explanatory variables will have a great impact on the model and most likely the results presented will be biased.

In need of a model that would reflect more accurately the behavior of economies, the conditional convergence hypothesis was introduced. Conditional convergence hypothesis takes into account a country's structural characteristics and predicts that countries will converge with one another after taking into account their structural differences (Tiiu, Kuusk, Schlitte, & Võrk, 2007). Under this concept countries that are poor relative to their own steady states would tend to grow more rapidly and convergence holds after controlling for differences in their steady states. This means that countries converge to their own steady states under a common theoretical framework and not to one common steady state (Weil, 2009). However the conditional convergence model fails to take into account countries initial conditions which are fundamental determinants of convergence.

In order to take both the initial conditions and the structural characteristics into account and create a more complete model, economists introduced the club convergence hypothesis. This hypothesis which also seems to best underline the empirical data implies that if countries or regions are similar in their structural characteristics and initial conditions, then, in the long run, will converge with each other in per capita income terms (Tiiu, Kuusk, Schlitte, & Võrk, 2007). However, it is important to mention here that similarity of countries in their initial conditions or structural characteristics is a rather subjective matter.

To be precise, if some countries have deviated a lot from the rest, this could be due to either the fact that these countries were not so similar with the rest at the first place (thus verifying the theory of convergence), or because they were similar and still did not converge (disregarding the theory of convergence).

To sum up, in the process of identifying convergence clubs with the use of empirical data, economists have been taking into account country specific initial conditions (some of which could be: labor force, capital stock, human capital, capital share, income per capita, savings etc.), the structural characteristics (such as: technology, services, population growth and density, unemployment, etc.) as well as the spatial dimension (geography plays a crucial role in identifying which clubs are formed and why) and policy reforms.

The Solow-Swan Growth Model

This section presents the mechanisms surrounding growth theory. As mentioned above, convergence theory (optimism) has its foundations in the neoclassical growth model that was originated by Solow and Swan, and states that the long run growth rate of output is determined by an exogenous rate of technological progress. In its basics the model is characterized by an aggregate production function with capital and labor being its main parameters (1), as well as a capital stock accumulation function (2).

$$Y_t = F(K_t, L_t) \qquad (1)$$

$$k_{t+1} = sf(k_t) - (n + \delta)k_t \qquad (2)$$

Where, Y_t is the output, L_t is labor, K_t is capital, s is the savings rate, n is the labor force growth rate and δ is the depreciation rate. The aggregate production function is homogeneous of degree 1, increasing, concave and twice continuously differentiable. In addition the model assumes that it has positive and diminishing marginal returns of factors inputs and constant returns to scale with respect to capital and labor. An extension of the above mentioned neoclassical growth model takes into account the technological progress as an exogenous factor (the AK model). In this model, the production and capital accumulation functions are adjusted as follows:

$$Y_t = F(K_t, A_t L_t) \qquad (3)$$

$$k_{t+1} = sf(k_t) - (n + g + \delta)k_t \qquad (4)$$

The equilibrium in the neoclassical growth model with exogenous technological progress is illustrated in the figure below.

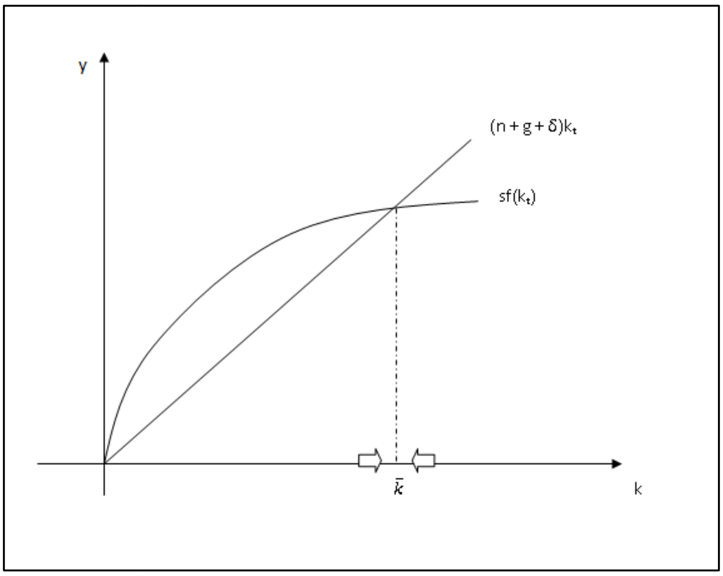

Figure 2: Neoclassical model equilibrium - with technological progress.
Source: Own illustration according to the Neoclassical Solow-Swan Growth model.

Convergence in the Solow-Swan model

The theory of convergence derives from Solow's neoclassical growth model with exogenous technological progress. The model consists of a Cobb-Douglas production function as presented in equation (5,) the steady state capital-labor ratio and the income per capita as given by equations (6) and (7) respectively:

$$Y_j = AK_j^a L_j^{1-a} \qquad (5)$$

$$\bar{k}_j = \left[\frac{sA}{n+\delta}\right]^{\frac{1}{1-a}} \qquad (6)$$

$$\bar{y}_j = A^{\frac{1}{1-a}} \left[\frac{s}{n+\delta}\right]^{\frac{a}{1-a}} \qquad (7)$$

From the above mentioned model we can infer that the only factors that can lead to differences in growth rates are either a varying level of technology or a varying growth rate of capital stock or labor force. However these are not the only influence channels that empirical literature suggests, and being a model that disregards distance and spatial discrepancies, the neoclassical growth model can be extended to better fit the empirics.

However, having understood the basics of growth and convergence, the parameters that can affect the steady state level of economies, are now more obvious. These parameters include, but are not limited to, the discount rate, depreciation rate, capital share, population growth rate and the inter-temporal elasticity of factor substitution. Other structural characteristics such as education, unemployment, services, socio-political characteristics capital accumulation and composition of sectors, can also be held accountable for differences in the steady states. Assuming that such parameters could be similar across economies, convergence in terms of income is expected (Kalyvitis, No Date).

Previous Empirical Literature

In the empirical literature economists have tried to analyze growth and convergence with the use of theoretical models such as the ones mentioned above. Analyzing the empirics of economic growth and convergence, in his theoretical model, Quah (1996) makes predictions on cross-section dynamics by assuming that countries select themselves into groups endogenously and thus do not act in isolation, that specialization in production allows exploiting economies of scale and that ideas are an important driver of growth. Quah presents two key finding regarding “coalitions” – club-convergence. He states that convergence clubs are formed endogenously and that different convergence clubs are formed depending on the initial distribution of characteristics across countries.

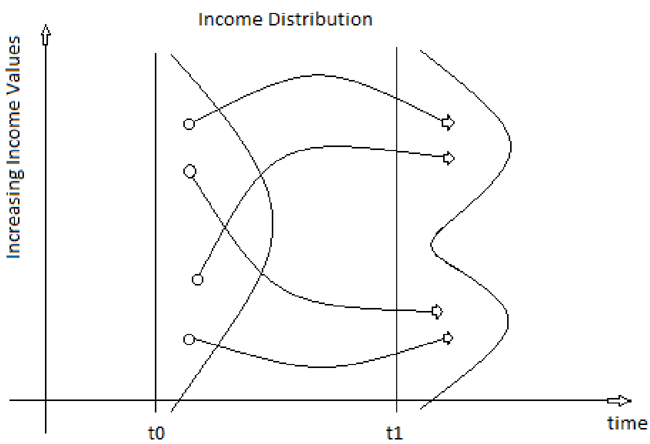

Figure 3: Evolving income distribution – tending to bimodal.
Source: Danny T. Quah; Empirics for economic growth and convergence.

As illustrated in the figure above, there is some initial income distribution across the cross section of economies at time zero, t_0. Over time, t_1, some economies evolve becoming better off, others fall behind and overtaking is possible.

Eventually the middle-income group vanishes while the rich become richer and the poor poorer. In this specific example, income evolution follows a bimodal distribution; however the exact outcome of clubs depends on the initial distribution of income among the entire set of countries. This means that if the initial income distribution is sparser at time t_0, then the evolution of income will result in multiple club formations (Quah, 1996).

Taking his research one step further, in another study Quah (1996) suggests that information externalities occurring at either the state or neighborhood level might be the cause of club-convergence. He suggests that nations that share borders or that belongs to geographically homogenous areas will tend to converge to the same club. In this way, Quah brings the spatial dimension into the picture and highlights its importance.

In a similar framework of thinking, Canova states that under the basic neoclassical growth model, production functions that illustrate decreasing returns to scale to the capital-labor ratio, exogenous population growth and fixed savings rate, may form convergence clubs under two specific occasions. First when savings rate out of wage and interest incomes differ and secondly when the economy features heterogeneous agents. The outcome in both situations will be the formation of multiple equilibria and club membership is determined by the distribution of the initial income per capita. Under such a framework, one can also consider additional elements such as human capital, fertility and market imperfections in order to suggest additional channels that support the club-convergence hypothesis (Canova, 2004).

With various views on the concept of club-convergence, the question of what generates it still remains. Summarizing the above mentioned views we can highlight the following channels through which clubs are formed. First of all indicators such as the initial levels of income, income distribution, human capital and human capital within the unit could be a good measure for explaining those heterogeneities that cause club convergence. In addition spatial factors need to be used in order to take into account the impact of neighborhood externalities and last but not least even policy variables can be used as a measure of national effects (Canova, 2004).

The importance of spatial dependence

This section underlines the importance of considering spatial dependence in any model of convergence and consequently in club-convergence. The process of economic growth and convergence is inherently endowed with a spatial dimension (Fischer & Stirböck, 2006). Spatial dependence is triggered by various factors such as labor and capital mobility, technology and knowledge diffusion, information spillovers, trade between countries, transportation and transaction costs, externalities as well as national factors. Anselin and Rey (1991), distinguish between two types of spatial dependence. The first type, substantive spatial dependence refers to the importance of externalities in growth theory. More specifically, knowledge spillovers arising from trade or commuting and migration flows (Fischer & Stumpner, 2008), are considered to have great contribution to economic growth and as a consequence on club convergence. The second type of dependence is referred to as nuisance spatial dependence and is caused by measurement problems connected to boundary mismatches or omitted variables.

In order to model spatial dependence, the regression equation has to be adapted to include either a spatially lagged dependent variable or a spatial error term (following a first order spatial autoregressive process), for substantive and nuisance spatial dependence respectively. However since this study of club-convergence takes place at the national level, the spatial effects are not considered to be as significant as in the case of regional analysis.

Ignoring spatial dependence can have grave consequences on the empirical analysis and the inferences made. In cases, where OLS (Ordinary Least Squares) estimation is used in linear regression analysis, spatial autocorrelation in the error terms violates the assumption of uncorrelated errors. This means that failing to take into account the spatial dimension will result in OLS estimates that will be biased, inefficient and misleading (Anselin & Rey, 1991).

The Impact of Economic Shocks

One of the major assumptions behind this research paper is that one of the factors that affect the steady state income level of a country is shocks which may vary from natural disasters and wars to economic and financial shocks. The current economic downturn which was the result of the collapse of the financial system in 2008 can definitely be categorized as such a shock that could throw an economy off its steady state and impact its growth. Being an economic shock, it is considered to be symmetric in the sense that its degree of impact does not vary depending on the geographical position of a country. To be more precise, geography cannot affect the nature or the impact of an economic shock.

When an economy goes into a recession, many macroeconomic indicators are negatively affected. For example, GDP (Gross Domestic Product), employment, investment spending, total factor productivity, household incomes, business profits, inflation, consumption and international trade, decrease. With all of these factors having a direct or indirect impact to both the steady state level and growth rate of a country, one would expect that there should be an impact on the club convergence patterns. In addition, macroeconomic policies that are taken in order to sustain growth and counterbalance the negative effects of the recession, should definitely affect the steady state level and growth rate of countries (Jones, 2002), which in turn affect the club convergence patterns.

The negative impact of the financial crisis of 2008 is obvious if we take a look at the data on GDP per capita obtained from EUROSTAT's database. As illustrated in Appendix 1 the average growth rate for the 27 member countries of the EU between 2007 and 2008 was 5%. With the hit of the financial crisis, all of the economies recessed experiencing zero or negative growth resulting in an average of -7.4% GDP per capita growth for the EU region. Taken as a whole, between 2008 and 2010, the EU's GDP per capita decreased by 3.7%. In a similar concept, final consumption per capita between 2008 and 2010 has decreased by 2% on average and unemployment for the same years has increased by 4.08%.

With such an impact on the economic performance and growth of countries in only three years, it is expected that convergence and club formation will be influenced to a certain extent. However, cautiousness is needed since convergence is a long-term process and its effects may come up in later years.

Empirical Modeling

Modeling Spatial Dependence

There are many methods to consider for identifying each country's club membership. One method would be to arbitrarily identify cut off levels of initial variables. This method however can easily produce biased results since club membership can vary depending on the cut off value chosen by the researcher. In a different approach, one can use spatially weighted test statistics of clustering or even formally test whether or not GDP per capita of countries move together with the use of time series data. In this report, with emphasis given to the spatial factor, spatially weighted test statistics will be used.

As mentioned before, various theoretical studies state that initial conditions are those that determine in which club each country belongs. However, with the same degree of importance, spatial dependences are also critical in determining club membership. Emphasizing those spatial differences, this research paper assumes that these are the main drivers of club formation. More specifically spatial differences in the initial level (beginning of the study period) of GDP per capita are thought to be the determinants of club formation.

Through the years researchers have used a number of techniques to test and measure spatial autocorrelation. These techniques vary from Moran's I statistic (commonly used among geographers) and Getis and Ord's G statistic to estimations of spatial autocorrelation coefficients of regression equations, used by spatial econometricians. As proposed however by the empirical literature, this paper will also use Moran's I-statistic to test for spatial autocorrelation and then employ the Getis and Ord's G-statistic to determine club membership.

The G-statistic introduced by Getis and Ord will be the main tool used to study club formation. The G-statistic is an index of global spatial autocorrelation for values that fall within a specific distance from each other. As pointed out by Getis and Ord, results indicate that the G statistics should be used together with I-statistics, so as to provide additional explanations as to those processes that give rise to spatial association. The values of Moran's I-statistic range from -1 to 1 indicating perfect dispersion or perfect correlation respectively. The G-statistic has the advantage of highlighting local "pockets" of dependence which would not show up with the use of other global statistics (Getis & Ord, 1992). The equation for Moran's I and Getis' and Ord's G statistics used, is presented below.

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (8)$$

Where, $S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}$

$$G_{it}^{*}(\delta) = \frac{\sum_{j=1}^{n} w_{ij}^{*}(\delta) y_{jt}}{\sum_{j=1}^{n} y_{jt}} \qquad (9)$$

In (8) and (9), y_{jt} represents the log-normal per capita GDP in country j at time t (t =1995), $w_{ij}^{*}(\delta)$is the (i, j) element of a row-standardized binary spatial weight matrix W* where $w_{ij}^{*} = 1$ if the distance from country i to country j (d_{ij}) is less than the critical distance δ, and $w_{ij}^{*} = 0$ otherwise.

By construction the elements of the main diagonal of W are set to zero to prelude an observation from directly predicting itself, however the GDP values of all countries (including the one for which the w_{ij}=0) are included in the y_{jt} matrix. For simplicity the standardized form of equation (9), presented below, is used in the analysis.

$$z[G_{it}^{*}(\delta)] = \frac{G_{it}^{*}(\delta) - E[G_{it}^{*}(\delta)]}{\sqrt{var[G_{it}^{*}(\delta)]}} \qquad (10)$$

In the above equation, positive values of $z[G_{it}^{*}(\delta)]$ represent spatial clustering of high values and negative values clustering of low values. This means that if a country has a positive $z[G_{it}^{*}(\delta)]$ it will be allocated to club 1, whereas if it has a negative one, it will be assigned to club 2.

Defining the threshold value of δ

One challenge that is presented at this point is how to define the value of δ. This value is critical to the outcome of club formation, since it represents the maximum distance between countries over which spatial influences have an effect. A too large or too small choice of δ can cause club misallocation. In a study conducted by Fischer & Stirböck (2006), where regional income convergence is studied, the authors choose this value a priori based on sensitivity analysis and theoretical considerations. They find the value of δ to be 350 kilometers.

Country	*Capital City*	*Average Distance from Neighboring Capital City (km)*
Austria	*Vienna*	347
Belgium	*Brussels*	319
Bulgaria	*Sofia*	409
Cyprus	*Nicosia*	915
Czech Republic	*Prague*	335
Denmark	*Copenhagen*	440
Estonia	*Tallin*	246
Finland	*Helsinki*	238
France	*Paris*	654
Germany	*Berlin*	577
Greece	*Athens*	831
Hungary	*Budapest*	347
Ireland	*Dublin*	463
Italy	*Rome*	819
Latvia	*Riga*	328
Lithuania	*Vilnius*	444
Luxembourg	*Luxembourg*	358
Malta	*Valletta*	688
Netherlands	*Amsterdam*	369
Poland	*Warsow*	554
Portugal	*Lisbon*	501
Romania	*Bucharest*	468
Slovakia	*Bratislava*	260
Slovenia	*Ljubljana*	380
Spain	*Madrid*	778
Sweden	*Stockholm*	577

Figure 4: Countries, Capital & Neighboring Distances.
Source: Illustration of own calculations in Excel.

To define a value for δ, the following framework of thinking was employed. Assuming that the capital cities attract the majority of a country's population and that they are the main source of economic activity, the absolute distance (in kilometers) between each capital city and the capital cities of neighboring countries was calculated. As neighboring countries were considered those that share a boarder with each other. Having found those distances, an average distance is first calculated for each country (figures are presented in the table to the right[6]) and then a total average of all the countries' distances. Consider for example Hungary whose neighboring countries are: Slovenia, Austria, Slovakia and Romania. So the average distance between Budapest and Ljubljana, Vienna, Bratislava and Bucharest, is found to be 347km. The average distance is found to be 482 km thus providing the author with a good approximation of the threshold value of δ.

[6] A web based calculator was used to measure the distance between the desired capital cities: http://distancecalculator.globefeed.com/World_Distance_Calculator.asp, last visited: 2011-08-14.

It is highly important to mention here that this method has three biases. The fact that simple averages are used in the estimation method above could be a source of bias. If the average distances between the capital cities of neighboring countries were weighted using some index (weights according to population of neighboring countries or even weights depending on their economic activity), the results could be different. The second bias is that by choosing the capital cities of countries we exclude the possibility of other cities being crucial economic centers of countries. For example in Germany we consider the distances of the neighboring capital only from Berlin whereas Munich is also a town with strong economic activity. The third bias stems from the concept of core vs. periphery countries. More specifically, for Germany (being in the center of Europe with many neighbors) it might not affect its club membership if Munich was chosen instead of Berlin, as an economic center. However for countries positioned in the periphery of the EU (i.e. Poland), a choice of an economic center closer or further away from the neighboring borders will have a greater impact in defining its club membership. Acknowledging the presence of those biases, they are left for further examination and the above-mentioned framework is still considered to capture a significant amount of the spatial forces at play.

Club Formations

Using the δ threshold value of 482 km, a binary spatial weight matrix can be created, whose values are either zeros or ones. The spatial matrix W as well as the vector of GDP per capita from 1995 are then inserted into Matlab in order to compute Moran's I and Ord's G statistics. The code for the computation of Moran's I was obtained from Matlab Central and was written by Felix Hebeler from the Geography department of Zurich University. The computations that are of interest to us are those for 1995 since we are focusing in spatial differences at the beginning of the study period. Moran's I for that year was estimated to have an average value of 0.5806. Analyzing the level of dependency among the observations, the positive I-statistic value of 0.5806 indicates a high degree of positive spatial autocorrelation among the GDP per capita values. Having underlined the presence of spatial autocorrelation, the code for the G-statistic is now run to compute the values of $G_{it}^{*}(482)$ and $z[G_{it}^{*}(482)]$. The results of these computations are presented in figure 5 below, whereas the Matlab M-file codes are presented in Appendix 2.1 and 2.2 respectively.

```
Command Window
New to MATLAB? Watch this Video, see Demos, or read Getting Started.

I =

    0.5806


z1995G =

  Columns 1 through 12

    0.1253    2.5439   -1.1340   -0.8595    0.5160    0.4095    0.2448    0.1964    1.8755    0.5419   -0.9047   -0.0071

  Columns 13 through 24

   -0.2557   -0.7239   -0.3429   -1.0759    1.3846   -0.9758    0.6420   -1.0759   -0.9241   -1.1340   -0.1362   -0.0975

  Columns 25 through 27

   -0.8369    0.2448    1.7592

fx >>
```

Figure 5: Results as shown in the Matlab command window.
Source: Matlab software screenshot.

Having verified the existence of positive spatial autocorrelation, the computed G-statistic values help filter GDP per capita and remove the spatial component that is embedded in the data. Columns 1 to 27 in figure 5 above contain the normalized G-statistic values of the 27 member countries of the EU, ranked alphabetically. If the value is positive then the respective country is categorized under the high income club and is the value is negative under the low income club. The result is the formation of the following two club pattern, as illustrated in the map of figure 6 below, where blue color is used for the high income club and green color for the low income club.

- **Club 1:** (12 Countries)
  - Austria, Belgium, Czech Republic, Denmark, Estonia, Finland, France, Germany, Luxembourg, Netherlands, Sweden, United Kingdom.
- **Club 2:** (15 Countries)
  - Bulgaria, Cyprus, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Malta, Poland, Portugal, Romania, Slovakia, Slovenia, Spain.

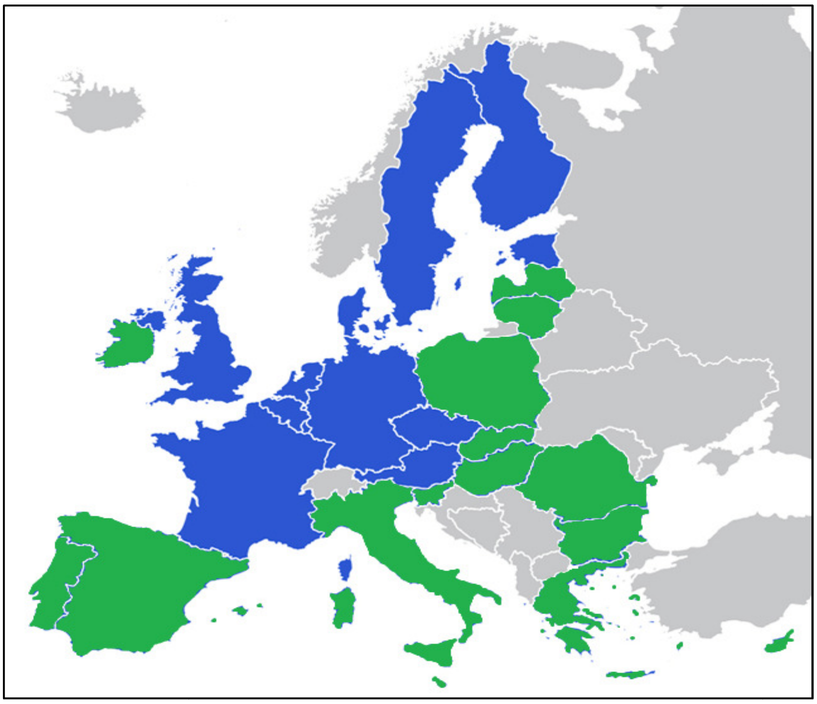

Figure 6: Geographical illustration of the two clubs formed.
Source: Own illustration according to the results obtained.

A first observation that is made when looking at the clubs formed is that high income countries tend to cluster together in one group and low income countries in another group. An exception to this trend is Estonia and Czech Republic, which despite their low GDP per capita level, cluster together with the rest of the high income countries in club 1. The simple explanation for this behavior could be that since these countries are neighbors with significantly high income countries, spatial factors may be at play and the influence might be really high. Last but not least, when calculating the G-statistic for the years 2008 or 2010 (Appendix 3) there is no significant change in the clubs formed. This illustrates the importance of both the geographical factor and the initial conditions. Those clubs that were defined in 1995 by their initial income per capita and taking into account the spatial factor, have not changed in 2008 and 2010, with the exception of Hungary, whose increase in GDP per capita and its proximity to Austria and Czech Republic, has led to a reallocation to club 1. The two club regime seems to be quite reasonable since the degree of geographic homogeneity within each club is quite high.

Beta Convergence Modeling

Among various methods for testing the convergence hypothesis, one of the most commonly used is beta-convergence analysis. Beta convergence which was first introduced by Baumol in 1986 is defined as a negative relationship between the initial income level and the subsequent income growth rate. Following the theoretical concepts put forth in the previous section, beta convergence is modeled differently depending on whether absolute or conditional convergence is tested. In its initial form, absolute beta convergence modeling is tested with the use of the following equation (Tiiu, Kuusk, Schlitte, & Võrk, 2007).

$$g_{j,t} = a + \beta y_{j,t} + \varepsilon_{j,t} \qquad (11)$$

Where, $\varepsilon \sim N(0, \sigma_\varepsilon^2 I)$.

In (11), g_T takes the form of $\frac{1}{T}\log\left(\frac{y_{i,t+T}}{y_t}\right)$ and represents the per capita GDP growth rate in period T, y_t represents the initial levels of GDP per capita, T is the length of the time period and ε represents the error terms. Parameters α and β are unknown. The model used in this research paper is that of conditional beta convergence. This model differs in the sense that negative correlation occurs only after taking into account each economy's structural characteristics. The cross-sectional equation used to test the conditional beta convergence is presented below.

$$g_{j,t} = a + \beta y_{j,t} + X\varphi + \varepsilon_{j,t} \qquad (12)$$

Where, $\varepsilon \sim N(0, \sigma_\varepsilon^2 I)$.

The additional term Xφ represents the set of explanatory variables and is introduced in order to control for the effect of structural differences. Beta convergence is identified when the value of β is significantly negative. In the empirical analysis of this research paper and being in line with the basics of conditional beta convergence modeling, the average national growth rate of an economy between two points in time (t_0 and t_0+T) will be given from the following equation.

$$\frac{1}{T}\log\left(\frac{y_{i,t_0+T}}{y_{i,t_0}}\right) = a + \beta\, y_{j,t_0} + \gamma X_{i,t_0} + u_{i,t_0} \qquad (13)$$

In the above equation (13), the term on the left hand side represents the economy's GDP per capita growth and T is the length of the time period. Analyzing the ring hand side of the equation, α can be interpreted as the equilibrium rate of GDP growth and β, which was introduced by Barro (1991) and Baro & Sala-i-Martin (1991), (1992), is used as a measure for computing the economy's convergence rate β*, given in the equation below.

$$\beta^* = -\frac{1}{T} ln(1 - T\beta) \qquad (14)$$

In equation (14), since T represents the time length of the study, the following restrictions are applicable: T≠0 and T≥1. In addition the contents of the parenthesis of the natural logarithm have to be positive, meaning that T*β<1. Equation 14 underlines the inverse relationship between an economy's growth rate and its steady state. From the above it is clear that as T increases β* decreases (Tselios, 2009). The graph below illustrates this inverse relationship.

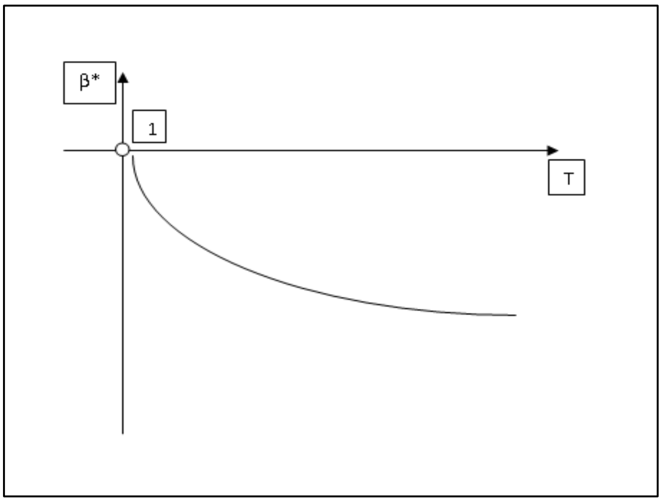

Figure 7: Graphical illustration of the beta coefficient.
Source: Own illustration according to equation 14.

Regression analysis of both the unconditional and conditional convergence models was performed with the use of the Ordinary Least Squares modeling and the use of cross-sectional fixed effects. Appendix 7 provides a detailed description of the basics of regression analysis as well as an insight into the statistics produced.

The choice of explanatory variables

In equation (13) which is used in the regression analysis, the term $\gamma X_{i,t_0}$ controls for the set of explanatory variables. These variables represent the structural characteristics of countries that could have an impact on growth and allow the model to test for conditional convergence. The question that arises here is which variables to choose and why. In order to answer this question, one has to go back to the basics of economic growth and find those factors that affect economic growth.

Taking a Keynesian approach to decide on which variables to choose, one would argue that investments, savings, government expenditure, R&D expenditure and education are some of the factors affecting economic growth. In a more market-based approach, macroeconomic stability, trade liberalization, capital mobility and exchange rate policies are those factors that influence economic development. Combining these two views with the evidence from the empirical literature regarding structural characteristics that affect growth, this paper considers the following variables as highly significant in affecting economic prosperity: *Unemployment, Educational attainment, Government Expenditure, R&D Expenditure, Active Population, Private Consumption and Balance of Exports-Imports.*

The choice of **unemployment** is based on the notion of Okun's Law. Okuns' law is mostly used as a rule of thumb and states that if GDP increases (decreases) by 2% above (below) the normal growth rate, then unemployment will decrease (increase) with 0.4%. This implies that the coefficient beta of GDP is expected to be negative. The above relationship has been tested by regressing GDP growth on the change in the unemployment rate (Blanchard, 2009). Running regressions using the opposite relationship would result in a negative coefficient, γ, for unemployment as well. The time series obtained on unemployment are expressed as a percentage of the total population. For that reason, when running regressions the term "*dlog*[7]" is inserted in front of the variable to account for the percentage change is unemployment.

Educational attainment was a rather straight forward choice. Empirical economic evidence supports the view that both public and private returns of investment to education have positive effects. Empirical studies show that if the upper level of schooling years is increased by 1 year, then the country's economic growth is also increased by approximately 0.44% per year

[7] $\text{dlog}(X) = d(\log(X)) = \log(X_t) - \log(X_{t-1})$

(Miller, 2008). In the empirical analysis, educational attainment is captured by measuring the percentage of the total population that has upper secondary and tertiary education. The γ coefficient of education is expected to be positive and highly significant. The use of educational attainment as an explanatory variable accounts for the effect of human capital on economic growth.

Even though there is evidence of great variation among countries, co-integration tests reveal that **government expenditure** is connected to the output of EU countries under a stable long run relationship. The common long-run elasticity is estimated to be positive (i.e. the γ coefficient is expected to be positive) and slightly below one in the long run. However this long-run relationship differs significantly over time and is more than one for catching-up, fast-ageing, low-debt and weak-rule countries. Country-specific short-term elasticity implies on average a speed of adjustment of government expenditure to potential output of about 3 years. (Arpaia & Turrini, 2008). Nonetheless in this paper, government expenditure is thought to be important in affecting economic growth and is measured as a percentage of GDP.

Another factor with great impact on GDP growth is **R&D expenditure**. Solow and Swan (1956), proved that technology and science are key factors for a country's economic growth. In addition, Romer and Lucas in the 1980s initiated the endogenous growth theory in which R&D is a crucial determinant of growth in the proposed economic model. Co-integration tests in the empirical literature prove the existence of a long-run positive relation between R&D expenditure and GDP growth. The parameter used in this empirical analysis is R&D expenditure expressed as a percentage of GDP per capita.

In the effort to capture all the potential determinants of economic growth, the contribution of a change in the **labor force** to economic growth, must also be taken into consideration. The first reason of such a consideration is that unemployment (also considered in the analysis) depends on the dynamics of labor demand and population[8]. The second reason is that growth

[8] The information was obtained from a website were important factors regarding unemployment are presented: http://www.economicswebinstitute.org/glossary/unemploy.htm, last visited: 2011-11-18.

in the labor force is a key factor that affects a nation's maximum sustainable or potential rate of economic expansion[9].

A growth in the labor force can act as a boost in an economy's growth but through the years, its contribution to GDP growth has illustrated diminishing returns. Empirical studies have shown that the channel through which labor force can impact economic growth is through employment. More specifically an increase in the labor force will result in an increase in the employment, which in turn has a positive relationship with GDP growth (Walterskirchen, 1999). For all the above mentioned reasons active population (as a percentage of the total population) was chosen as a variable to measure labor force.

A parameter that can have a huge impact on the growth of GDP per capita is the amount consumed by individuals. **Consumption** is the largest component of the GDP[10] and thus small changes in it can have a significant impact on GDP per capita and in turn economic growth. It represents the actual and imputed expenditure of households and is mainly a measure of both goods and services consumed by individuals. Following the definition of consumption, this paper uses *household final consumption expenditure* as a measure of private consumption (as obtained from EUROSTAT database). This measure accounts for expenditure incurred by resident households on goods and services.

Within the European Union one of the factors that promote growth is international trade and trade liberalization. Even though the relation of trade barriers to GDP growth is more complicated,[11] empirical literature suggests that **export and import** shares of GDP are significantly and positively correlated with growth. This is mainly due to the fact that trade promotes economic growth through a number of channels such as technology transfers and economies of scale. As a result, the net balance of exports and imports (as a percentage of GDP) is used in the regression analysis and its coefficient is expected to be positive and highly significant.

[9]Federal Reserve Bank of San Francisco on the determinants of labor force growth: http://www.frbsf.org/publications/economics/letter/2007/el2007-33.html, last visited: 2011-11-18.
[10] From the expenditure approach of GDP calculation: GDP = C + I + G + (X-M).
[11] Trade barriers are found to be positively and significantly correlated with economic growth, especially for developing countries (Yanikkaya, 2003).

Empirical Results & Analysis

Testing unconditional convergence

A crucial part of the empirical analysis is to understand how the conditional convergence hypothesis differs from the unconditional one. To be precise, in order to understand the dynamics of the model it would be very useful to see how the model of unconditional convergence changes, after taking into consideration those structural characteristics mentioned in previous sections. For that reason, regression analysis starts by first testing the unconditional model as presented in equation (11). The exact regression equation takes the following form:

$$GDPGRW = C + \beta \log\left(GDP(-1)\right) + \varepsilon_{j,t} \qquad (15)$$

In the left hand side of equation (15), GDP growth was chosen as the dependent, rather than the alternative of $\frac{1}{T}\log\left(\frac{y_{i,t+T}}{y_t}\right)$, in order to avoid having to deal with simultaneity bias. In the right hand side of the regression equation, the predictor variable (GDP) is put on a log scale. With the introduction of the logarithm, the intercept, C, represents the expected value of the response if GDP is 1 and the slope measures the expected change in GDP growth, when GDP is increased by a fixed percent. In addition the logarithm helps in smoothing out extreme fluctuations in the data. A one year lag has been introduced, accounting in this way for the assumption that previous year's GDP affects this year's growth. This equation is then used for running regression in both clubs, using cross-section fixed effects.

High Income Club

Testing first for the years up to 2008 and then including the years of economic turbulence (up to 2010) a summary table has been created, using the outputs generated by the EViews statistical software. Figure 8 below presents these findings.

	Up to 2008				Up to 2010			
Variables	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
C	0.469528	0.259997	1.805895	0.073	0.954472	0.413397	2.308853	0.0222
LOG(GDP(-1))	-0.041578	0.026345	-1.578198	0.1167	-0.090536	0.04191	-2.160247	0.0322
R^2	0.476786				0.361717			
Adjusted R^2	0.43288				0.315853			
S.E. Of regression	0.04027				0.050378			
Akaike criterion	-3.506749				-3.069041			
Schwartz criterion	-3.252594				-2.838438			
Durbin-Watson stat.	1.54979				1.721458			

Figure 8: Summary of regression results of unconditional convergence.
Source: Own illustration according to Eviews output.

Interpretation of the results before the economic crisis illustrated above shows that unconditional convergence is valid within the high income club countries. More specifically using the measurement coefficient of LOG(GDP(-1)), β=-0.0415, into equation (14), will generate a value for β^* = -0.0332. This means that the average rate of convergence for the years up to 2008 was 3.32%. However the significance of the β coefficient is not high, showing that a careful look in the statistics is necessary. The model's explanatory power is quite weak (low value for the R^2 and adjusted R^2). In addition, the Durbin-Watson statistic is lower than 2, which may be an indication of serial autocorrelation in the residuals. Providing a further insight into the statistics, EViews is used to perform likelihood tests, the results of which are presented below.

Redundant Fixed Effects Tests
Equation: REG2
Test cross-section fixed effects

Effects Test	Statistic	d.f.	Prob.
Cross-section F	3.344524	(11,143)	0.0004
Cross-section Chi-square	35.715191	11	0.0002

Figure 9: F- & Chi-square statistics (up to 2008).
Source: EViews output.

Figure 9 above shows the estimated values for the Chi-square statistic[12]. This statistical test measures whether or not there exists a relationship between the dependent variable (GDP growth) and the explanatory variable, in our case GDP per capita.

[12] Test of the null hypothesis of a non existing relationship between the dependent and the predictor variables, against the alternative hypothesis of an existing relationship.

The large number of the Chi-square statistic, combined with the low probability value points to the rejection of the null hypothesis and verifying that the GDP per capita of the previous period affects the growth of this period. Last but not least, one of the main assumptions for the choice of a linear model was that of the normality of the error distribution. Figure 10 below was generated after executing a normality test on the residual series of the estimated regression equation (15). The histogram verifies the assumption of a bell-shaped error term distribution.

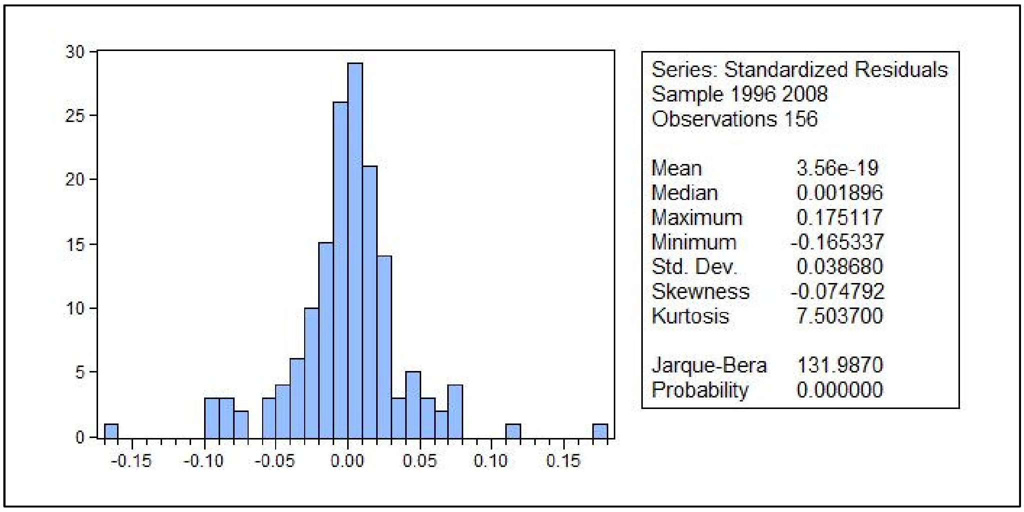

Figure 10: Normality test of the residual series (up to 2008).
Source: EViews outputs.

The question that arises here is how does unconditional convergence behaves if the analysis is extended to account for the years following the financial crisis of 2008. In order to be able to capture potential changes, the same regression equation is estimated under the same specifications as before, but this time with the time interval extending to 2010. The right hand side of the summary table of figure 8 above presents the results of the second set of regression analysis.

What is evident here is that convergence is present and this time highly significant at the 95% confidence level. In detail, using the estimated β coefficient of LOG(GDP(-1)) in equation (14) as before, will provide a value of $\beta^*=-0.0572$, implying that the average rate of unconditional convergence within the high income countries is 5.72%, after taking into account the “economic turbulence” years. An in-depth look into the summary statistics shows that R-squared and adjusted R-squared are quite low and even lower than before, meaning that lots of the variation in GDP growth cannot only be explained by the previous period’s GDP.

Contributing to this, the standard error of the regression has increased pointing out that the economic downturn is affecting the model by increasing its error. Including 2009 and 2010 in the analysis however, has resulted in an increased value for the Durbin-Watson statistic, bringing the value closer to 2 and decreasing the presence of serial autocorrelation in the residuals (if any).

Figures 11 & 12 below represent the likelihood and normality tests respectively, this time for the time series up to 2010. The highly significant and large value of the Chi-square statistic illustrates that despite adding the crisis years, the null hypothesis of a no relationship between GDP growth and the previous period's GDP can be rejected. In addition, normality in the error term distribution is still evident in figure 12, continuing to support the concept of a linear regression model.

Redundant Fixed Effects Tests
Equation: REG2
Test cross-section fixed effects

Effects Test	Statistic	d.f.	Prob.
Cross-section F	2.678116	(11,167)	0.0034
Cross-section Chi-square	29.243043	11	0.0021

Figure 11: F- & Chi-square statistics (up to 2010).
Source: EViews output.

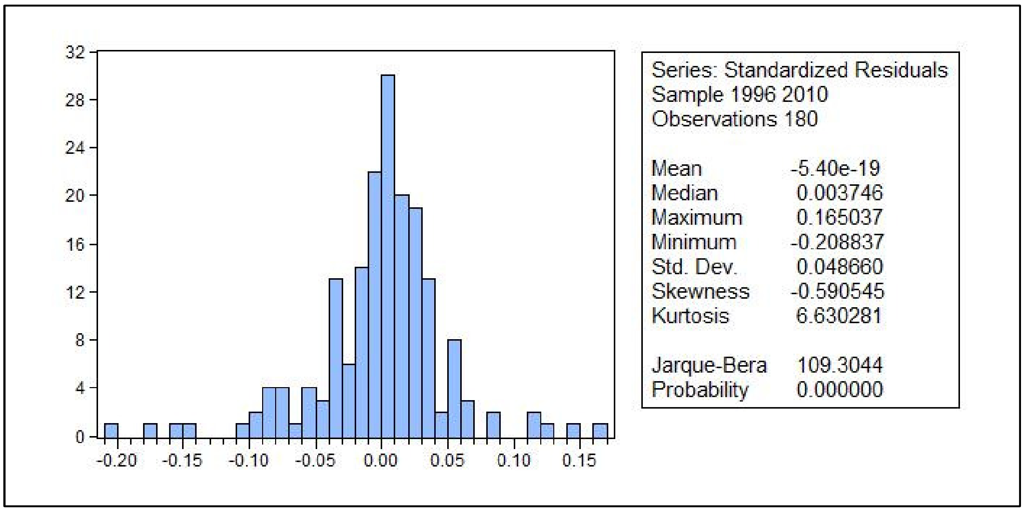

Figure 12: Normality test of the residual series (up to 2010).
Source: EViews output.

Low Income Club

In line with the regression analysis performed before, the focus is now turned towards the low income club countries. This section differs however in the fact that the regression equation used is in its exact form as equation (12), without considering a logarithmic factor.

$$GDPGRW = C + \beta GDP(-1) + \varepsilon_{j,t} \qquad (16)$$

The decision to exclude the logarithm is attributed to the fact that the GDP per capita of club 2 countries does not illustrate any extreme fluctuations and thus smoothing out could be omitted. The first step is to test the regression equation for the years up to 2008. The results are presented below.

	Up to 2008 without the AR(1) term			
Variables	***Coefficient***	***Std.Error***	***t-Statistic***	***Probability***
C	0.127619	0.019735	6.466482	0.0001
LOG(GDP(-1))	-0.0084	0.0041	-2.034694	0.0434
R^2	0.302577			
Adjusted R^2	0.244134			
S.E. Of regression	0.065598			
Akaike criterion	-2.532047			
Schwartz criterion	-2.263493			
Durbin-Watson stat.	1.608446			

Figure 13: Regression results of unconditional convergence (up to 2008). Source: Eviews output.

Unconditional convergence is verified for the low income club as well. The β coefficient of GDP can be used in equation (14) and calculate a β* rate of 0.8%. As shown by the t-statistic and its low probability value, this result is highly significant at the 95% confidence interval. However as before, with only one explanatory variable to support GDP growth, the model lacks explanatory power (low values of R^2 and adjusted R^2). To add to that, the Durbin-Watson statistic of 1.6084 (significantly lower than 2) suggests that serial autocorrelation in the residuals is present.

As mentioned before, serial autocorrelation can produced biased statistics for the model and is order to account for it a first order autoregressive term can be introduced[13]. Adding the AR(1) terms results in a model with no serial autocorrelation and a higher explanatory power. In addition the standard error of the regression is lower than before. The new β coefficient of GDP is -0.0096 and gives an increased β* rate of -0.91%, which is still highly significant. A closer look at the summary statistics shows that the Akaike of the AR(1) model is higher, making the model a better fit to the data. Figures 14 and 15 present the results of the likelihood and normality tests respectively.

Redundant Fixed Effects Tests
Equation: REG1
Test cross-section fixed effects

Effects Test	Statistic	d.f.	Prob.
Cross-section F	1.942930	(14,163)	0.0254
Cross-section Chi-square	27.779631	14	0.0152

Figure 14: F- & Chi-square statistics of the ar(1) model (up to 2008).
Source: EViews output.

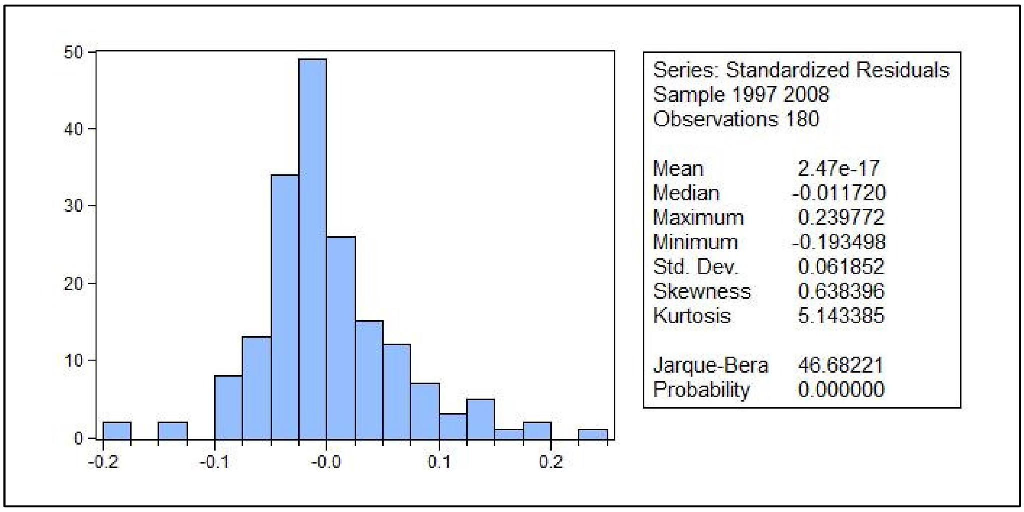

Figure 15: Normality test of the residual series, including ar(1) (up to 2008).
Source: EViews outputs.

The Chi-square statistic of 27.77 is significant at the 95% confidence level, rejecting in that way the null hypothesis that the dependent and predictor variables are not related. In addition the normal distribution of the error terms (depicted in figure 15) verifies the linear regression model's assumption of normality in the error terms.

[13] The first order autoregressive term takes the following form: $X_t = c + \varphi X_{t-1} + \varepsilon_t$.

Evidently the first order autoregressive model seems to produce more robust results and will be used as the basis of analysis for the low income countries. A summary of the results generated by EViews, after the inclusion of the AR(1) terms, is presented below, accompanied by the results of the likelihood test.

	Up to 2008					*Up to 2010*			
Variables	***Coefficient***	***Std.Error***	***t-Statistic***	***Probability***		***Coefficient***	***Std.Error***	***t-Statistic***	***Probability***
C	0.135514	0.019106	7.092717	0.0001		0.190199	0.022545	8.436311	0.0001
LOG(GDP(-1))	-0.0096	0.0041	-2.364851	0.0192		-0.0241	0.0045	-5.317362	0.0001
AR(1)	0.132258	0.069584	1.900689	0.0591		0.168085	0.068148	2.46647	0.0145
R^2	0.376903					0.317464			
Adjusted R^2	0.31574					0.260881			
S.E. Of regression	0.059106					0.071192			
Akaike criterion	-2.729298					-2.369386			
Schwartz criterion	-2.427741					-2.09843			
Durbin-Watson stat.	1.842389					1.89067			

Figure 16: Summary of the regression results of unconditional convergence with AR(1) process. Source: Own illustration according to Eviews output.

Redundant Fixed Effects Tests
Equation: REG1
Test cross-section fixed effects

Effects Test	Statistic	d.f.	Prob.
Cross-section F	1.688721	(14,193)	0.0605
Cross-section Chi-square	24.266859	14	0.0425

Figure 17: F- & Chi-square statistics of the ar(1) model (up to 2010). Source: EViews output.

The average convergence rate within the low income club has increased to -2.08% after taking into account the years of economic turmoil. The impact of the economic downturn on the variables has resulted in lower values for the R^2, adjusted R^2 and the Akaike information criterion, meaning that the model's explanatory power is reduced. Adding to that, the standard error of the regression has increased to 0.071. Such a behavior of the statistics is consistent however with modeling shocks and is anticipated.

Inferences and conclusions

What it is observed within the unconditional convergence framework, after the introduction of the economic turbulence years, is that the average rate of convergence has increased for both the high and the low income club. This contradicts the hypothesis tested in this research paper, namely that the economic recession should have a negative impact on the rate of convergence within a club of countries. Considering GDP per capita of the previous period as the only explanatory variable of growth has generated a relatively weak model and the probability still remains high, that the impact of the economic downturn on each country's structural characteristics may strongly affect the clubs average convergence rate when incorporating the crisis.

Another reason that may have cause the unconditional convergence rate to increase despite the economic downturn is that most of the countries in both clubs have either experienced growth or stabilization in GDP per capita during 2010. This means that the decreased values of GDP per capita of 2009 were offset by a similar or larger increase during 2010. To add to the list of explanations of why unconditional convergence rate has not slowed down after the crisis, comes the fact that convergence in its nature is a long run process. To be precise, the effects of the economic instability, which is far from over and still up to today negatively affects countries' economies, may be more visible in some years from now.

A last possible explanation of the observations within the unconditional convergence analysis presented above is related to the steady state level of countries. Early in this research paper, it was mentioned that there will not be any analysis targeted to whether or not each country's economy is at a steady state level. Even though it is not accurate, one could assume that the majority of the developed economies of the high income club (Club 1) are relative stable in terms of population growth, consumption, etc. and thus quite close to their respective steady states. Such assumptions however cannot be made for the low income club (Club 2) in which most of the economies are quite unstable and presumably far from their respective steady states. Assuming that these countries are not at their steady state levels, it implies that the crisis has acted as a catalyst, by increasing convergence within club two but most likely towards a lower level steady state.

Even though the explanations above seem valid, further examination is required within the concept of the conditional convergence hypothesis. In this way, it will be easy to check which of these still hold under a more detailed framework and which are rejected. The next section contains the analysis and interpretation of the results within the conditional convergence framework.

Testing Conditional Convergence

The unconditional convergence framework that was tested in the previous section has provided some strong foundation for understanding the mechanisms behind convergence within the two clubs. However it has not succeeded in answering all the questions surrounding club convergence and has initiated an increased interest for further analysis. This section focuses in testing the conditional convergence framework, whose specification is underlined by equation (12). In more detail, substituting the variables in use, equation (12) takes the following form.

$$GDPGRW = C + \beta\, GDP(-1) + \gamma X_{i,t_0} + u_{i,t_0} \qquad (17)$$

Where, u_{i,t_0}represents the error is term and $\gamma X_{i,t_0}$the set of explanatory variables representing the structural characteristics of countries. It would be wise to see a break-down of this term in all of the explanatory variables.

$$\gamma X_{i,t_0} = \gamma_1 UNEMP + \gamma_2 EDU + \gamma_3 R_D + \gamma_4 LABOR$$
$$+\gamma_5 CONS + \gamma_6 GOV + \gamma_7 EXIMP \qquad (18)$$

The combination of equations (17) & (18) constitute the model used in regression analysis of the conditional convergence framework. Following similar steps as before, this equation will be first tested for the high income club and then for the low income club. The coefficient of GDP per capita is expected to be negative and highly significant, whereas all of the γ coefficients (excluding unemployment) are expected to be positive with varying significance depending on the time interval and club studied. To be more precise, when testing within club conditional convergence for the whole time interval (including the economic instability years up to 2010), some of the predictor variable are expected to lose their significance due to the effects of the economic crisis. In addition, for the countries of club 2, being quite different in their structure not all of the predictor variables are expected to be significant (i.e. R&D expenditure might not be so important for GDP growth).

High Income Club

Within the high income club regressions are performed with the use of all of the variables representing structural characteristics, since they are all considered to be significant in influencing GDP growth. Following the same structure as before, the results for both before and after the crisis period are presented in figure 18 below.

	Up to 2008					Up to 2010			
Variables	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*		*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
C	0.555348	0.188639	2.943972	0.0038		0.296686	0.239124	1.240723	0.2165
LOG(GDP(-1))	-0.038786	0.013729	-2.825167	0.0054		-0.028282	0.014378	-1.96697	0.0509
DLOG(UNEMP)	-0.037048	0.014431	-2.567228	0.0113		-0.048948	0.017446	-2.805728	0.0056
LOG(EDU)	0.069885	0.027505	2.539553	0.0122		0.030806	0.033703	0.914056	0.3621
LOG(R_D(-1))	0.032542	0.013002	2.502799	0.0135		0.021164	0.017666	1.198007	0.2327
EXIMP	0.267934	0.066336	4.039025	0.0001		0.213903	0.054103	3.953637	0.0001
GOV(-1)	0.2333	0.090688	2.572568	0.0112		0.210872	0.076959	2.740037	0.0068
LOG(LABOR)	0.175166	0.057931	3.023695	0.003		0.037348	0.081845	0.456324	0.6488
DLOG(CONS)	0.977261	0.042454	23.01947	0.0001		1.04883	0.051953	20.18819	0.0001
R^2	0.918509					0.911906			
Adjusted R^2	0.907125					0.901445			
S.E. Of regression	0.016297					0.019121			
Akaike criterion	-5.276507					-4.971644			
Schwartz criterion	-4.8855					-4.616871			
Durbin-Watson stat.	2.132069					1.998217			

Figure 18: Summary of regression results of conditional convergence testing.
Source: Own illustration according to EViews output.

The regression output presented above supports the expectations set regarding the behavior of the predictor variables. For the time series ending in 2008, the β coefficient of LOG(GDP(-1)) is negative and highly significant at the 95% confidence interval, implying an average rate of convergence of β*=3.14% for the high income club. All of the other explanatory variables are also highly significant at the 95% confidence interval and their respective signs reflect the existing expectations.

In detail, unemployment has a negative sign implying that a percentage increase in unemployment by 1 would result in a 3.8% decrease in GDP growth. Education is positively related to economic growth showing that increased investments in secondary and tertiary education of 1% result in almost 7% of economic growth. In total there is a positive relation between GDP growth and R&D expenditure, net exports, government expenditure (with 1 year lag), active population and consumption.

All of the t-statistics are significantly higher than 1.96 indicating that all of these variables are relevant in the model. The model's adjusted R^2 is 0.907, suggesting that approximately 90% of the variation in GDP growth can be explained by the predictor variables.

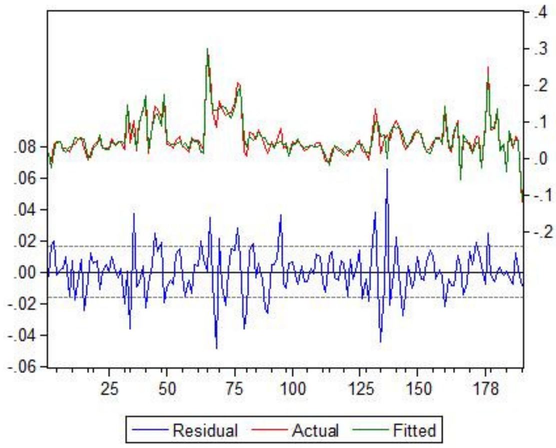

Figure 19: Actual-Fitted-Residual graph (2008). Source: EViews output.

As illustrated in the graph on the right-hand side, the equation (fitted: green line) performs really well in predicting the actual data (actual: red line). Also shown in this graph is the residual series. Residuals are fluctuating mostly within the desired interval of ±2 and as pointed out by the Durbin-Watson statistic of 2.1 the residuals are not containing any serial autocorrelation. A last look at the summary statistics shows that the standard error of the regression is considerably low and the absolute values of the Akaike and Schwartz criterion are high enough to suggest that the regression model is a relevant in explaining the data.

To be certain regarding the robustness of the results, likelihood and normality tests are performed, the results of which is presented in figures 20 and 21 respectively. The significantly large value of the Chi-square statistic suggests that there is a relationship between economic growth and the set of explanatory variables, whereas the bell-shaped distribution of the residuals verifies the use of a linear regression model.

Redundant Fixed Effects Tests
Equation: REG1
Test cross-section fixed effects

Effects Test	Statistic	d.f.	Prob.
Cross-section F	2.744039	(11,136)	0.0031
Cross-section Chi-square	31.269153	11	0.0010

Figure 20: Likelihood test for the conditional convergence model (up to 2008). Source: EViews output.

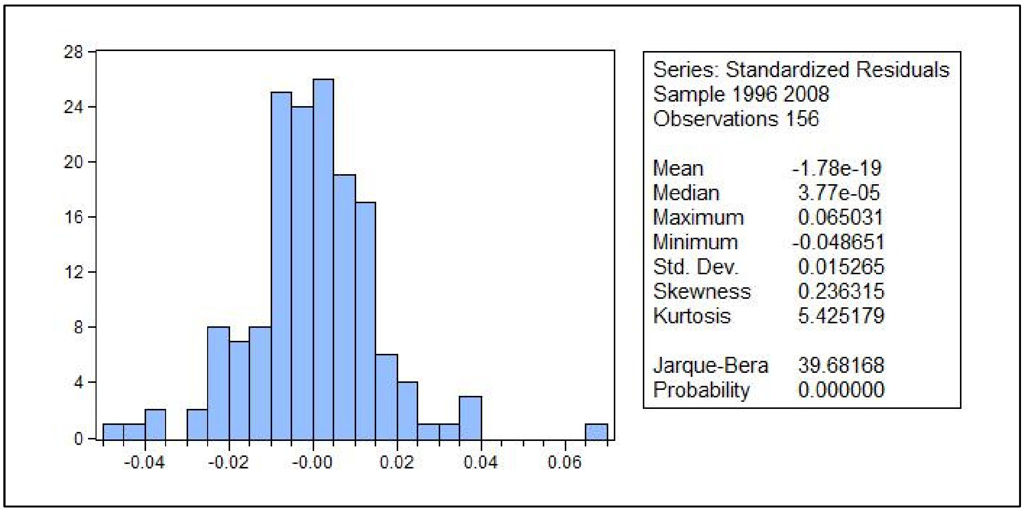

Figure 21: Normality test on the residuals (up to 2008).
Source: EViews output.

The results of figure 18 illustrate that net exports, government expenditure and consumption have the largest impact on economic growth and the main reason for this is that these predictor variables are the ground components of GDP itself. Excluding consumption from the estimated model would result in an increased impact (β coefficients) from the rest of the variables, increased probabilities and a new confidence interval of 90%, increased standard error for the regression, decreased R^2 and adjusted R^2, decreased Akaike and Schwarz and a Durbin-Watson significantly lower than 2, suggesting serial autocorrelation in the residuals. All the above changes indicate that the initial model (including consumption) responds better in explaining economic growth and conditional convergence.

The right hand side of the table presented in figure 18 above illustrates the results of the regression analysis after extending the time series to 2010. Comparing the regression output with that of the previous time interval provides a good insight in the channels and the impact of the economic crisis. Starting from GDP per capita, the β coefficient has decreased to -0.0283 which results in an average rate of convergence of 2.36%, among the countries of club 1. The new β coefficient is still highly significant but this time at the 90% confidence level, which is still acceptable considering the large amount of data analyzed. An examination of the coefficients of the structural characteristics demonstrates that in terms of effect (positive vs. negative), the results are consistent with the expectations set in the theoretical part of the report. However, when it come to their significance, not all are significant at the 95% or 90% confidence level. More specifically, education, R&D expenditure as well as active population are no longer significant for the model and their respective t-statistics are way below the 1.96 threshold value.

To add to that when the coefficients are compared before and after the crisis period, significant changes are observed. The negative effect of unemployment on economic growth has increased with the impact of the financial crisis. Moving at the same direction, consumption's impact on growth has also increased and is the biggest compared to all other variables. This increase might be an attempt to counter-balance the decrease in the impact of the rest of the predictor variables.

The summary statistics of the regression model illustrate that despite the effect of the economic shock, the model still performs well. The R^2 and adjusted R^2 have slightly decreased whereas the standard error has increased. These changes are expected since the years that were added in the sample incorporate great economic instability. The Durbin-Watson statistic is almost 2 meaning that there isn't any serial autocorrelation in the residuals that could bias the statistical estimates of the OLS. Generally as illustrated by the figure on the right, the tested regression equation succeeds in predicting the path of economic growth.

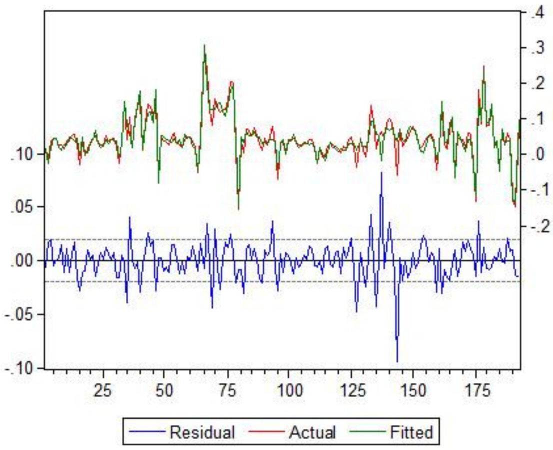

Figure 22: Actual-Fitted-Residual graph (2010).
Source: EViews output.

As a last test of robustness, a likelihood test is performed, indicating the Chi-square statistics. The values presented in figure 23 below are evident of a model where there is strong relationship between economic growth and the set of explanatory variables chosen.

Redundant Fixed Effects Tests
Equation: REG1
Test cross-section fixed effects

Effects Test	Statistic	d.f.	Prob.
Cross-section F	2.273524	(11,160)	0.0133
Cross-section Chi-square	26.141285	11	0.0062

Figure 23: Likelihood test for the conditional convergence model (up to 2010).
Source: EViews output.

Inferences and conclusions for Club 1

Having performed regression analysis both before and after the financial crisis makes it possible to compare the results from the two points in time and make some inferences regarding the effects of the economic downturn on club convergence and its mechanisms. Starting the analysis from the β coefficients of the models it is obvious that $\beta_{2010} < \beta_{2008}$and consequently $\beta^*_{2010} < \beta^*_{2008}$. The average rate of convergence has decreased from 3.14% for the year ended in 2008 to 2.36% for the year ended in 2010, in absolute terms. This shows that after incorporating the years when the economies where hit by a shock, within club convergence has slowed down.

However it is highly important to mention here that at this point in time, it cannot be proven (empirically) whether the slowdown in convergence is solely attributed to the economic downturn. Recall that the question of convergence asks if the differences among economies are getting smaller over time. This means that a decreased convergence rate may be the cause of economies reaching a common steady state and an evidence of minimization of their differences. On the other hand, with economies experiencing different population growth rates, different consumption and different incomes, it would be unrealistic to assume that economies have approached so close to a common steady state.

Last but not least, answering to the question of whether or not this slow down in convergence will persist or even increase over time; one could argue that such an outcome depends on the economy's ability to recover after the crisis. As mentioned before, the European economy is still highly unstable and convergence is a long-run process.

Shifting the analysis toward the set of structural characteristics, there have been two major observations regarding their estimated coefficients for the whole time span. The first observation was that certain variables became insignificant as soon as the "crisis" period was taken into account. These variables where: educational attainment, R&D expenditure and active population. Theoretical and empirical evidence suggest that the educational sector is directly impacted in the case of a recession. This negative impact is the outcome of increased fees, cut-downs of investments in education, lay-offs and reduced budgets[14].

[14]Article summarizing the issues facing the education sector: http://libcom.org/library/recession-what-it-means-education, last visited: 2011-11-23.

As a result educational attainment (at the upper secondary and tertiary levels) impact on growth becomes smaller and insignificant. R&D expenditure is also impacted during a recession with most of the firms in the economy cutting down the expenses on innovations and putting more emphasis in savings (Yunlu & Murphy). Last but not least, active population is no longer significant due to the fact that unemployment has increased making the percentage of people available to "produce" insignificant since they remain unemployed. These views on the effects of an economic recession, justify the observed insignificance in certain variables.

The second observation regarding the structural characteristics is that their effect on economic growth, that is the β coefficients, has changed with the impact of the crisis. In general terms the effect of the variables having a negative impact on economic growth has increased, while the effect of those variables that contribute to GDP growth has decreased when the economy is hit by the crisis. This is in line with the assumptions regarding the effects of an economic recession. An exception to that rule is consumption whose effect, despite the crisis, has increased. When considering the variable of consumption, it is used in the regression with a "dlog" in front of it, representing thus the percentage increase in consumption from the previous year. In addition, as mention before, consumption is the largest component of GDP and as a result it is correlated with the variables of GDP, net exports and government expenditure. The decreases in those correlated variables are partly counterbalanced by an increased consumption coefficient, implying that during a recession, consumption plays a very important role.

If regression analysis did not include the variable of consumption, what would be observed is a model with increased effects from the coefficients part but with much lower R^2 and adjusted R^2, reduced Akaike and Schwartz values, increased standard error of the regression and a Durbin-Watson statistic indicating the presence of serial autocorrelation. In total the estimated model without consumption performs poorly compared to the one before and this is also illustrated in figure 24 to the right.

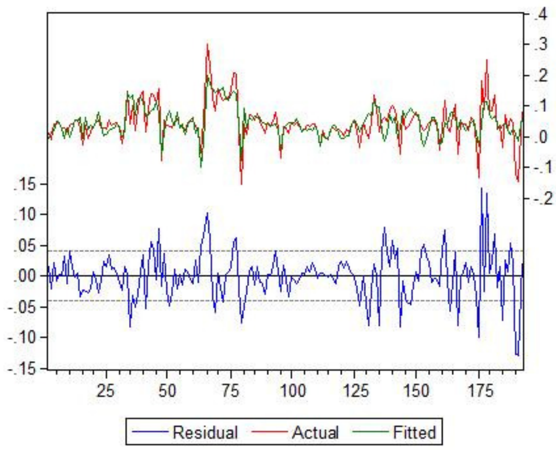

Figure 24: Actual-Fitted-Residual graph (no cons.). Source: EViews output.

Low Income Club

Before diving into the empirical analysis of the low income club, it is important to set some expectations regarding the variables and behavior of those countries. The first thing that it is important to mention here is that for many variables, usually closer to the beginning of the time period of study, there were missing data on most of the countries. This issue was overcome by using time period average growths in estimating the omitted variables. However this resulted in only estimating average values rather than the actual path and could act as a source of bias. Having in mind the behavior of the high income countries, it is quite accurate to say that the structural characteristics of the low income club members will most likely not behave in the same way. This is due to the fact that in reality, in terms of their structural characteristics the low income countries are very different from the high income countries. Despite their membership in the European Union, these differences endure over time. In this linear model, estimated with the least squares method, certain explanatory variables are not expected to behave so well in terms of significance, statistical sign and even result in biases

In the empirical analysis of within the low income club, not all parameters illustrated in equation (18) were included in the model. To be exact, the variable of government expenditure was left outside the model. The reason for this decision was that during the regression analysis government expenditure was not only found highly insignificant but with a negative coefficient, contradicting expectations regarding the relationship between economic growth and government expenditure. In addition the model's summary statistics were found to be more robust when government expenditure was excluded. This behavior of government expenditure verifies the expectations mentioned before. The results of the regression analysis within the low income club countries are presented below in figure 25.

	Up to 2008 with the AR(1) term			
Variables	***Coefficient***	***Std.Error***	***t-Statistic***	***Probability***
C	0.009878	0.019111	0.516874	0.606
GDP(-1)	-0.0066	0.0062	-3.901044	0.0001
LOG(UNEMP)	-0.016296	0.006799	-2.396615	0.0177
LOG(EDU)	0.041607	0.024807	1.677222	0.0955
DLOG(R_D)	0.030342	0.042354	0.716404	0.4748
EXIMP	0.133202	0.063795	2.087972	0.0946
DLOG(LABOR)	0.179442	0.106709	1.68161	0.0946
DLOG(CONS)	1.020352	0.063085	16.17432	0.0001
AR(1)	-0.332367	0.090901	-3.65638	0.0003
R^2	0.868934			
Adjusted R^2	0.850568			
S.E. Of regression	0.027621			
Akaike criterion	-4.221635			
Schwartz criterion	-3.813646			
Durbin-Watson stat.	1.931775			

Figure 25: Regression results of conditional convergence testing of club 2 (up to 2008).
Source: Own illustration according to EViews output.

The results of the estimation show that for the years before the financial crisis (ending in 2008), convergence is present but at a much lower rate than in club 1. The β coefficient of GDP takes a value of -0.0066 implying an average rate of conditional convergence of β*=0.64%. Nonetheless the β coefficient is highly significant at the 95% confidence interval. The analysis of the set of explanatory variables shows that the signs of their coefficients behave as expected but significance varies. For most of the predictor variables significance lies at the 95% or 90% confidence interval. R&D expenditure however is an exception to this trend. The percentage change in R&D expenditure is found to be insignificant in affecting economic growth, illustrating a t-statistic of 0.71 (significantly lower than 1.96). The first order autoregressive terms is used again to correct for serial autocorrelation in the residuals.

As a whole the model performs really well in predicting the behavior of GDP growth. The high values of R^2 and adjusted R^2 suggest that the model succeeds in explaining approximately 85% of the variation in GDP growth. The standard error of the regression equation seems to be reasonably low while the Durbin-Watson statistic of 1.93 rejects any questions of serial autocorrelation in the residuals.

Figure 26 on the right illustrates how well equation (17) fits the actual data as well as the variation of the residuals. However when tests are run to determine the robustness of the model and its estimates, the results seem to suggest that the statistic produced by the model, even though they are accurate, might be misleading. The Chi-square values of the likelihood test are insignificant at the 85% confidence level, suggesting that one or more of the explanatory variables (or the way they are used) are not having a direct relationship with the GDP growth.

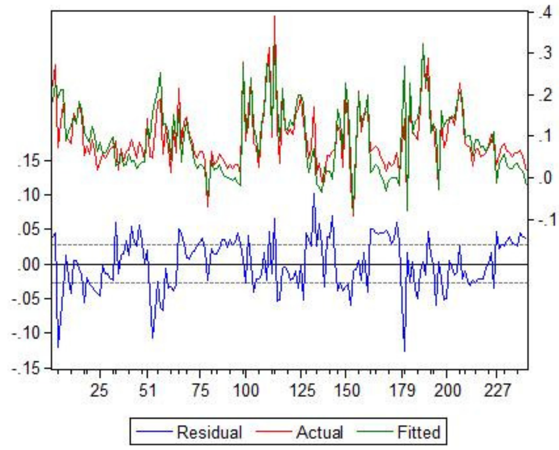

Figure 26: Actual-Fitted-Residual graph (2008). Source: EViews output.

Redundant Fixed Effects Tests
Equation: Untitled
Test cross-section fixed effects

Effects Test	Statistic	d.f.	Prob.
Cross-section F	1.266301	(14,157)	0.2341
Cross-section Chi-square	19.257460	14	0.1553

Figure 27: Likelihood test for the conditional convergence model of club 2 (up to 2008). Source: EViews output.

Continuing with tests on robustness, the normality test shows that the error distribution is still bell-shaped and the Jargue-Bera value verifies this at the 95% significance level[15]. Nonetheless there is a trend toward bimodal distribution pattern as illustrated by figure 28 below. This could be evidence of a low degree of heterogeneity present in the model which could increase after extending the time interval. It is important to point out that in the presence of heterogeneity, the residuals cannot be used in any statistical tests since the results will be misleading.

[15] To understand the model's significance, the Jargue-Bera values should be compared with the Chi-square table with two degrees of freedom.

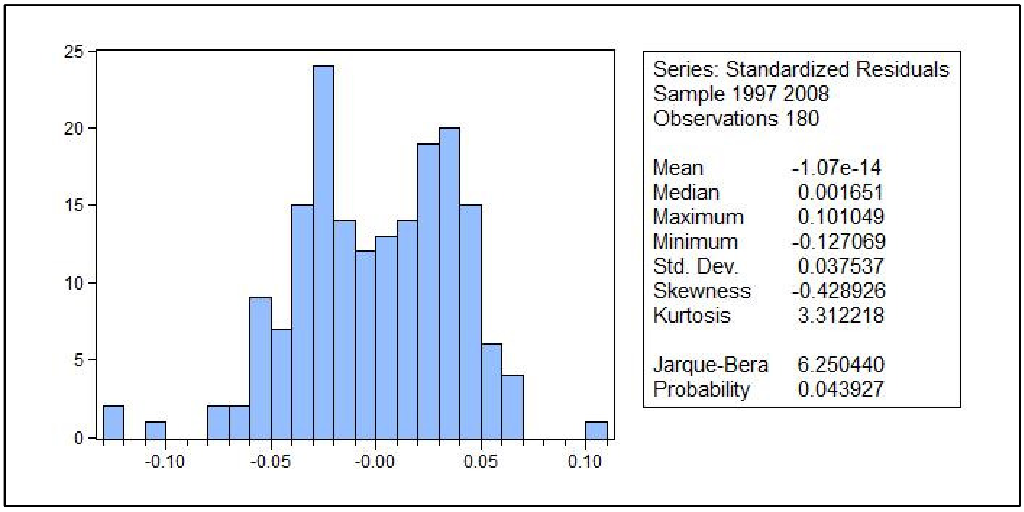

Figure 28: Normality test on the residuals (up to 2008).
Source: EViews output.

In an effort to understand the source of this potential heterogeneity in the model, the regression equation is estimated again but this time without the variable of consumption. The new robustness tests illustrate a highly significant relationship between the set of explanatory variables and economic growth, and a persistent non-normality of the error term distribution increases. In addition the first order autoregressive term becomes insignificant and if it is excluded from the model then the coefficient of net exports becomes negative and insignificant. In an overview of the model, R^2 and adjusted R^2 have decreased to 44% and 36% respectively, the standard error of the regression has increased to 0.056 and Akaike and Schwarz criterion absolute values have decreased to 2.78 and 2.39 respectively.

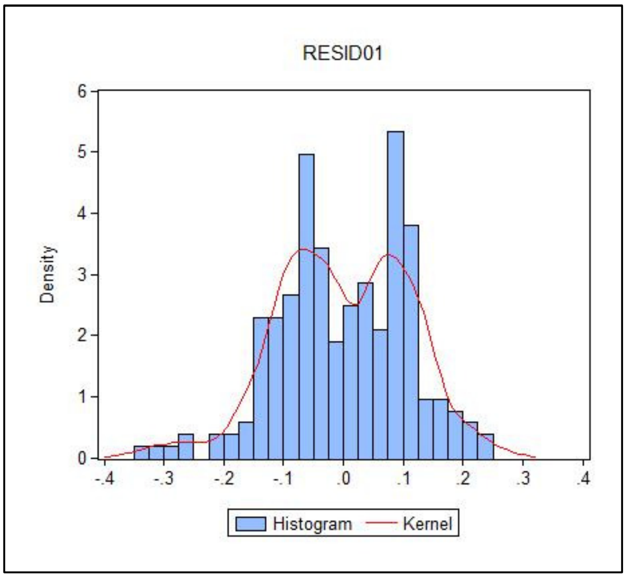

Figure 29: Normality test on the residuals (up to 2010).
Source: EViews output.

Considering all the above and in fear of heterogeneity present in the model, normality tests are performed for the whole time period. Figure 29 verifies the presence of heterogeneity and as a consequence the above model cannot be used.

The conclusion of the above analysis was that the variables of net exports and consumption, being correlated to GDP per capita, created serial autocorrelation in the error terms. When the first order autoregressive scheme is used, serial autocorrelation is dealt with but on the other hand the appearance of heterogeneity poses a problem. Such an issue could be solved with the use of a method different from least squares estimation. However since least squares estimation is used throughout the report for reasons of consistency and comparison, the decision was taken to exclude both the explanatory variables that are correlated with GDP per capita and the first order autoregressive term. Figure 30 below, presents a summary of the regression results before and after the economic crisis.

	Up to 2008					***Up to 2010***			
Variables	***Coefficient***	***Std.Error***	***t-Statistic***	***Probability***		***Coefficient***	***Std.Error***	***t-Statistic***	***Probability***
C	-0.387837	0.095401	-4.065333	0.0001		-0.386542	0.10974	-3.522354	0.0005
GDP(-1)	-0.008	0.0049	-5.928465	0.0001		-0.01	0.0071	-5.219805	0.0001
LOG(UNEMP)	-0.059702	0.022193	-2.690122	0.0078		-0.082585	0.021081	-3.91758	0.0001
EDU	0.682023	0.157828	4.321302	0.0001		0.598466	0.141213	4.238033	0.0001
DLOG(R_D)	0.135605	0.089638	1.512804	0.1321		0.14899	0.076996	1.935028	0.0544
DLOG(LABOR)	0.461143	0.270013	1.707853	0.0894		0.372073	0.28627	1.299725	0.1952
R^2	0.406425					0.376561			
Adjusted R^2	0.34198					0.318779			
S.E. Of regression	0.061205					0.070226			
Akaike criterion	-2.65225					-2.38952			
Schwartz criterion	-2.316557					-2.085867			
Durbin-Watson stat.	1.829891					1.720494			

Figure 30: Summary of the regression results of conditional convergence testing of club 2.
Source: Own illustration according to EViews output.

The new model still supports the hypothesis of convergence within the low income club. The β coefficient is estimated to be -0.008 which when used in equation (14) results in an average rate of conditional convergence, β*, of 0.76. The coefficients of unemployment and education are positively significant at the 1% level and so is the coefficient of active population but at the 10% probability level. R&D expenditure, even though it has a positive coefficient, it is not significant. Such a behavior of the R&D variable is expected since low income countries do not engage in a R&D activities.

The summary statistics of the model suggest that it doesn't perform so well in explaining the behavior of economic growth. R^2 and adjusted R^2 illustrate that only 40% of the variation in GDP growth is explained by the set of explanatory variables. This is also reflected in the figure to the right. The standard error of the regression model is relatively low, 0.06, and so are the Akaike and Schwatz criterion values. Nonetheless the Durbin-Watson statistic verifies the absence of serial autocorrelation in the residuals and as shown by the likelihood and normality tests below, the new model is much more robust.

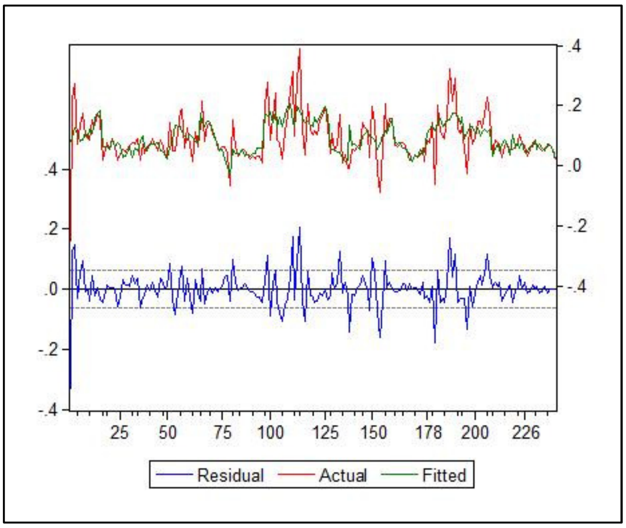

Figure 31: Actual-Fitted-Residual graph (2008). Source: EViews output.

Redundant Fixed Effects Tests
Equation: Untitled
Test cross-section fixed effects

Effects Test	Statistic	d.f.	Prob.
Cross-section F	3.353345	(14,175)	0.0001
Cross-section Chi-square	46.342113	14	0.0000

Figure 32: Likelihood test for the conditional convergence model of club 2 (up to 2008). Source: EViews output.

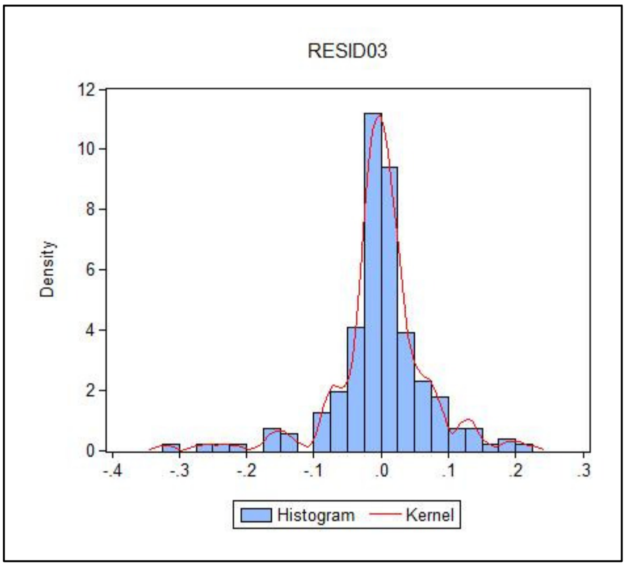

Figure 33: Normality test on the residuals (up to 2008). Source: EViews output.

The Chi-square value produced by the likelihood test is quite large and highly significant at the 1% probability level. This confirms the presence of a strong relationship between economic growth and the set of predictor variables used. In addition the normality test validates the use of a linear model and is a strong indication of homoskedasticity in the model. Having found an appropriate and robust model to test for in the low income club, the next step is to perform regression analysis but this time after taking into account the economic "turbulence" years up to 2010.

Taking a second look at the summary table under figure 30, it is clear that conditional convergence within the low income club is still present and has also increased despite of the effects of the crisis. In more detail, the β coefficient of GDP has increased to -0.01, implying an average rate of convergence, β*, of 0.93. The next section labeled "Inferences and Conclusions of Club 2" tries to analyze this behavior of convergence within the low income club. The signs of the coefficients of the predictor variables remain as before and are still significant at either the 1% or 10% probability level. An interesting change however is that R&D expenditure is now significant at the 90% confidence interval whereas active population has become insignificant. The summary statistics of the regression's output show a reduced value for the R^2 and adjusted R^2 as well as an increased value of the regression's standard error. This is something anticipated when modeling periods of economic instability. In line with the above, the Akaike and Schwartz criterion values have also slightly decreased. The Durbin-Watson statistic of 1.72 could be interpreted as an indication of serial autocorrelation in the residuals but in such a large sample with 225 observations, such a value is acceptable.

Before going into the comparison of two periods (before and after the crisis) however, the same robustness tests as before are performed. The Chi-square large value of 53.21 with a very low probability level suggests that even after incorporating the years of economic instability, there is a strong relationship between the dependent and independent variables. Not illustrated here, the Jargue-Bera high values and significance as well as the bell-shaped error distribution, verify the use and validity of the linear model.

Inferences and conclusions for Club 2

When testing the impact of the economic downturn on the conditional convergence rate of the low income countries, the results of the regression analysis tell a totally different story compared to the findings of the high income club. Starting the interpretation of the results with the β coefficients of the two samples (before and after the crisis), there is no evidence of a slowdown in the average convergence rate. What is observed is that $\beta_{2010} > \beta_{2008}$ and as a consequence $\beta^*_{2010} > \beta^*_{2008}$. As with the unconditional convergence framework the average rate of convergence has increased and this time even the inclusion of structural characteristics did not influence the behavior of the β coefficient. However, comparing the absolute rate of convergence itself between the unconditional and conditional frameworks, it is obvious that after the inclusion of the predictor variables, β*, has decreased to a more reasonable level. The increase in the convergence rate means that countries move faster toward a common level in terms of output level and similarities in their structural characteristics. However such a common level does not have to be equated to higher growth or higher steady state level. Nevertheless the inferences on possible explanations for this kind of behavior are now even more valid that before.

As illustrated in figure 38 of the Appendix 5, most of the countries composing the low income club have actually rebounded during 2010. The huge plunge of GDP per capita in 2009 was offset by an increase or stabilization in 2010. As mentioned before, this might be one of the factors explaining the fact that convergence rate did not slow down and the persistent results of the conditional convergence model give additional support to this. But this alone is not enough; the most important fact affecting the conditional club convergence rate is the current situation of countries, meaning the proximity to their respective steady states. Assuming that countries within the low income club are far from their steady states, the result above shows that the crisis might have acted as a catalyst in bringing the low income countries together but it makes no inferences regarding the new level of the economies after the crisis (which is probably lower).

Taking the analysis one step further, one of the reasons that could explain this unexpected behavior of the convergence rate within the low income countries, is the level of functionality of these economies before the hit of the economic crisis.

It is possible that some of these economies were overwhelmed by high unemployment rates, decreased investments in education and R&D expenditure, weak institutions and failed policy reforms even before 2008. Their relatively "dysfunctional" economic states might have acted as an airbag absorbing the extreme negative effects of the economic crisis. If this is the case, the negative effects of the economic downturn on the low income club will be visible in the years to come, explaining why at this point an increase in the convergence rate is monitored. Speculating about the countries' economic rebound under this concept it is fair to say that the process of economic stabilization and growth in the future will probably be quite slow. With convergence being a long-run process, this matter is left for further examination in the future. Addressing the issue of whether such convergence behavior will be persistent over time, it is wise to say that the outcome of convergence depends on the path of the economies after the crisis and as a result, having investigated only two years of economic instability, it is too soon to say.

A very important observation is that while in the high income club, the impact of the financial crisis on the countries' structural characteristics was high enough to cause a slowdown in the convergence rate, in the low income club, the crisis impact on the countries' characteristics, even though visible and significant, was not so strong to cause a slowdown in the convergence rate. Examining the way the structural characteristics where affected reveals that the negative impact of unemployment has increased. This could be attributed to the fact that within country unemployment has increased in the last two years and as a result any additional increases in unemployment have a larger negative impact on growth. Related to the behavior of unemployment, the variable for active population went from being significant in the years leading up to the crisis, to being insignificant in 2010. In a similar view as in the high income club the percentage of the people available to work does not strongly influence a country's economic growth if quite a large portion of the population is unemployed. The economic recession had a negative impact on the coefficient of educational attainment as well. The β coefficient decreased from 0.68 to 0.59. This can be explained by the fact that the educational sector is highly impacted during an economic downturn period through increased tuition fees, reduced investments and lay-offs.

The discussion of the variable of R&D expenditure was left for last since it is similar to the behavior of GDP per capita in that it does not behave as expected. The coefficient of R&D expenditure which was insignificant in the pre-crisis period increased in magnitude and became significant in the post-crisis period. A possible explanation for this behavior of the variable of R&D is that as presented by empirical findings from time-series analysis of GDP and R&D expenditure, it is possible for investments in R&D to increase during periods of economic downturn (European Comission, 2011). Other research papers have illustrated that countries and companies that maintain or increase their R&D spending during periods of economic instability, tend to perform better and survive thought the crisis. The increase of the coefficient of R&D expenditure and its high significance could largely reflect the fact that low income countries (on average) might have made an effort to maintain their R&D investment levels.

This part concludes the analysis of the conditional convergence hypothesis within the low income club. The fact that a slowdown in convergence is not verified for the low income club leaves plenty of space for further testing and analysis. In addition the instability of the initial model which included consumption, net exports and government expenditure suggest that the use of another estimation method (other than LS) might be more appropriate for the low income club. However, this instability in the linear model reflects the importance of considering other factors, other than politics, when making assumptions of similarities among countries and considering club or even European Union memberships.

Conclusions

The first aim of this research paper was to understand the mechanisms of club convergence by highlighting the importance of initial conditions and spatial dependences. The findings suggest that the initial position of a country (in terms of output level) is highly important in determining club membership in the bimodal club convergence model. Equally important however is each country's position in the geographical space and its distance from the neighboring countries. The model used to account for the impact of the spatial factors seems to be performing well in defining the two clubs as well as country membership. The results suggest the formation of a high income club with 12 country members and a low income club with 15 country members. However, the use of this model does not prohibit the use of either simpler models based only on whether or not countries share a border, or even more complicated ones defining club formation with the use of temperatures.

The second and main objective of the paper was to study the effects of the economic crisis on the convergence rates within each club formed. The results obtained from the empirical testing shed significant light on the mechanisms of convergence. The high income club of countries seems to perform relatively better in terms of the linear equation tested with least squares model. The findings within the unconditional convergence framework illustrated that convergence still increases despite the negative impact of the economic crisis. On the other hand, as soon as the structural characteristics are included and conditional convergence is tested, the high income club countries experience a slowdown in the convergence rate when the years of 2009 and 2010 are incorporated to the time series of study. This provides strong evidence for the assumption of the crisis having a negative impact on the within club convergence rate. Even though this evident slowdown may be attributed to other factors (such as proximity to the steady state level) it seems rather unlikely.

The empirical analysis of the low income countries should have taken a straight forward path as before, however the analysis of the exact same model as the one used for the high income countries was surrounded by many difficulties. The nature of the data obtained (estimated missing values for the first years of the sample, for most of the countries) combined with the fact that countries belonging to the low income club tend to differ significantly in terms of structure, growth, institutions, etc., contributed to a model that was simpler in that it contained less explanatory values.

In spite of the fact that the model used was simpler the findings of the regression analysis where especially interesting. They suggest that the average rate of convergence between the countries of the low income club has increased despite the impact of the economic crisis. This result is supported by both the unconditional and conditional convergence framework implying that the structural characteristics did not affect convergence in the same way as in the high income club, either not being significantly impacted by the crisis or being positively affected by it (as in the case of R&D expenditure).

In the discussion of whether these trends will persist over time the inferences are the same for both clubs, suggesting that at this point in time it is too early to provide an accurate answer. The current economic instability of most of the countries within the low income club could definitely have an extreme negative impact on both clubs decreasing the rate of convergence further, in the future. However, the scenario of a temporary impact of the economic shock is also possible provided that the countries recover from the crisis in the years to come.

The description of the relation between GDP growth and the set of explanatory variables used seems to be strongly supported by the linear regression model used. The method used for the regression analysis, namely ordinary least squares, performs quite well for the high income club of countries and the statistical results indicate a high level of robustness. On the other hand, the fact that this model had to be simplified in order to be used for the empirical analysis of the low income club of countries provides an opportunity for additional testing under another method that could take into account more factors. To be precise, previous empirical literature at the regional level has suggested that a maximum likelihood estimation model that takes into account the spatial dimension tend to perform better than the OLS estimation.

Evaluating the model at the national level produces quite accurate results however it is believed that a national model that considers spatial dependence might be more accurate. As mentioned before, one of the restrictions that lead to the choice of a study at the national level was that regional data were available only up to 2008.

The unexpected findings within the low income club suggest that a similar study at either the regional or national level but with a longer time series (in the aftermath of the financial crisis) could give additional insight into the effects of the financial crisis on convergence within clubs. Last but not least, not examining the relative position of the two clubs to each other before and after the crisis allows for further discussion and empirical testing.

To sum up, the financial crisis of 2008 has resulted in an economic crisis which has impacted the countries' economies at a great extend. The research paper highlights those impacts and at the same time successfully points out the channels and type of influence on the within club convergence hypothesis.

Appendix

1. EU Countries

	Year of Entry	*GDP per capita 1995*	*GDP per capita 2007*	*GDP per capita 2008*	*GDP per capita 2009*	*GDP per capita 2010*	*% increase 2007-2008*	*% increase 2008-2009*	*% increase 2009-2010*	*% increase 2008-2010*
Austria	1995	23000	33000	33900	32900	34100	2.72%	-2.94%	3.64%	0.58%
Belgium	1952	21500	31600	32299	31500	32600	2.21%	-2.47%	3.49%	0.93%
Bulgaria	2007	1200	4000	4600	4600	4800	15.0%	0.0%	4.34%	4.34%
Cyprus	2004	10900	20300	21800	21200	21700	7.38%	-2.75%	2.35%	-0.45%
Czech Republic	2004	4100	12300	14200	13100	13800	15.44%	-7.74%	5.34%	-2.81%
Denmark	1973	26600	41700	42500	40300	42200	1.91%	-5.17%	4.71%	-0.7%
Estonia	2004	2000	12000	12200	10300	10700	1.66%	-15.57%	3.88%	-12.29%
Finland	1995	19600	34000	34900	32500	33600	2.64%	-6.87%	3.38%	-3.72%
France	1952	20200	29600	30100	29300	29800	1.68%	-2.65%	1.7%	-0.99%
Germany	1952	23600	29500	30100	29000	30300	2.03%	-3.65%	4.48%	0.66%
Greece	1981	9500	20300	21100	20800	20400	3.94%	-1.42%	-1.92%	-3.31%
Hungary	2004	3400	9900	10500	9100	9700	6.06%	-13.33%	6.59%	-7.61%
Ireland	1973	14400	43500	40500	35900	34900	-6.89%	-11.35%	-2.78%	-13.82%
Italy	1952	15100	26000	26200	25200	25600	0.76%	-3.81%	1.58%	-2.29%
Latvia	2004	1500	9200	10100	8200	8000	9.78%	-18.81%	-2.43%	-20.79%
Lithuania	2004	1400	8500	9600	7900	8300	12.94%	-17.7%	5.06%	-13.54%
Luxembourg	1952	38700	78100	81200	76600	82100	3.96%	-5.66%	7.18%	1.1%
Malta	2004	7300	13300	14200	14100	14800	6.76%	-0.7%	4.96%	4.22%
Netherlands	1952	20700	34900	36200	34600	35400	3.72%	-4.41%	2.31%	-2.2%
Poland	2004	2800	8200	9500	8100	9300	15.85%	-14.73%	14.81%	-2.1%
Portugal	1986	8900	16000	16200	15900	16200	1.25%	-1.85%	1.88%	0.0%
Romania	2007	1200	5800	6500	5500	5700	12.06%	-15.38%	3.63%	-12.3%
Slovakia	2004	2800	10200	11900	11600	12100	16.66%	-2.52%	4.31%	1.68%
Slovenia	2004	8100	17100	18400	17300	17300	7.6%	-5.97%	0.0%	-5.97%
Spain	1986	11600	23500	23900	22900	23100	1.7%	-4.18%	0.87%	-3.34%
Sweden	1995	22000	36900	36100	31300	37000	-2.16%	-13.29%	18.21%	2.49%
UK	1973	15200	33700	29600	25300	27400	-12.16%	-14.52%	8.3%	-7.43%

Figure 34: Analyzing the development of GDP per capita.
Source: Own illustration in Excel according to the data obtained from EUROSTAT.

2. MATLAB code

2.1. Code for the I-statistic

```
function M = morans_I(grid,W,s)
% PURPOSE: calculate global Moran's I for an input grid (matrix) by
% calculating all local Moran's I for a given moving windows
% size using a %weight matrix.
%-----------------------------------------------------------------
% USAGE: M = moransI(grid, W, s);
% where: [grid] is the matrix to analyse
% [W] is the normalized weight matrix of the size the local Moran's
% I will be calculated for (uneven sized!)
% [s] is an optional flag to use zscores of input values for
%    calculation. Set to 'true' if zscores of local grid should be
%    calculated. Leave blank if not desired or input values are already
%    standardized.
% -------------------------------------------------------------------
% OUTPUTS: [M] matrix of all local Moran's I
% -----------------------------------------------------------------
% NOTES: Weight matrix needs to be 'moving window' style, not contiguity
%       matrix: Moran's I is calculated and weighted for neighbours to %       center cell.
%       Matrix needs to be normalized (weights sum to 1) and center cell weight
%       will be set to 0 if not already. Uses localmoran.m
%       -> Use nanmean(M(:)) to get the average global Moran's I.
%
% See Anselin (1995, 'LISA.', Geogr. Analysis 27(2),p.93f) for details on
% standardized variables in calculation of local Moran's I.
%
% EXAMPLE:  M = moransI(rand(20,20),ones(5,5),'true')
%
% Felix Hebeler, Geography Dept.,de University Zurich, March 2006.

%% Check if standardising should be done
if exist('s','var')
   if strcmp(s,'true');
      grid=zscore(grid);
   elseif strcmp(s,'false')
      %do nothing
   else
      error('Invalid option for s: set [true] to calculated zscores to determine local Moran or leave blank if values are already standardized.');
   end
end
if (mod(size(W,1),2)| mod(size(W,2),2))~=1
  error('Weight matrix W needs to have uneven size (eg. 5x5)')
end
%% Do local Morans I calc of the grid.
M = NaN(size(grid,1),size(grid,2));
wsx=floor(size(W,1)/2);
wsy=floor(size(W,2)/2);
% Do local morans I calc for moving window ws
for row=1+wsy:1:size(grid,1)-wsy;
   for col=1+wsx:1:size(grid,2)-wsx;
      M(row,col) = get_moran(grid(row-wsx:row+wsx,col-wsy:col+wsy),W);
   end
end

%% calculate local Moran's I
```

```
function m=get_moran(raster,W)
ncols= size(raster,2);
nrows= size(raster,1);
zi = raster(ceil(nrows/2),ceil(ncols/2));%  value of center cell (note: no weight applied!)
if (isnan(zi));
   m=NaN;
   return;
end;
raster=raster.* W; % Weight values in window
raster(ceil(nrows/2),ceil(ncols/2))=0; %set center cell to zero to exclude zi from sum
zj = nansum(raster(:)); % sum of weighted values excluding zi
m = zi * zj; % calculate local Moran's I and return
```

2.2. Code for the G-statistic

```
% This matlab code computes the G statistic introduced by Getis and Ord, as
% well as Moran's I statistic. The results are shown in Matlab's comman
% window.
% Author: Johannes H. Stefanoudakis
% Date: 2001-10-18
% Number of spatial units indexed by i and j in 1995.
n=27;
% Computing S0:
for i=1:27
   for j=1:27
      r(i,j)=sum(W(i,j));
      K=sum(r);
   end
end
S=sum(K);

% Computing Morna's I statistic:
for i=1:27
   for j=1:27
      g(i,j)=(W(i,j)*(Y(i)-mean(Y(:,1)))*(Y(j)-mean(Y(:,1))));
   end
   h(i)=((Y(i)-mean(Y(:,1)))^2);
end
G=sum(g);
Gsum= sum(G);
H=sum(h);
I=(n/S)*(Gsum/H)

% Computing Getis and Ord G statistic:
for i=1:27
   for j=1:27
      k1995(j)=W(i,j)*Y1995(j);
      k2008(j)=W(i,j)*Y2008(j);
      k2010(j)=W(i,j)*Y2010(j);
      G1995(i)=(sum(k1995))/(sum(Y1995));
      G2008(i)=(sum(k2008))/(sum(Y2008));
      G2010(i)=(sum(k2010))/(sum(Y2010));
   end
end

z1995G(i)=(G1995(i)-mean(G1995))/(sqrt(var(G1995)))
z2008G(i)=(G2008(i)-mean(G2008))/(sqrt(var(G2008)));
z2010G(i)=(G2010(i)-mean(G2010))/(sqrt(var(G2010)));
```

3. G-statistic results and Club formation

	G_{1995}	zG_{1995}	$Club_{1995}$	G_{2008}	zG_{2008}	$Club_{2008}$	G_{2010}	zG_{2010}	$Club_{2010}$
Austria	0.1227	0.1253	1	0.135	0.323	1	0.1357	0.3252	1
Belgium	0.3447	2.5438	1	0.318	2.5629	1	0.3234	2.5947	1
Bulgaria	0.0071	-1.1339	1	0.0168	-1.1231	1	0.0163	-1.1179	1
Cyprus	0.0323	-0.8595	1	0.0331	-0.9242	1	0.0338	-0.9066	1
Czech Republic	0.1586	0.516	1	0.1368	0.3453	1	0.1408	0.3875	1
Denmark	0.1488	0.4094	1	0.1102	0.02	1	0.1131	0.0517	1
Estonia	0.1337	0.2448	1	0.1417	0.4048	1	0.1393	0.3686	1
Finland	0.1292	0.1963	1	0.1263	0.217	1	0.1268	0.2177	1
France	0.2834	1.8754	1	0.2631	1.89	1	0.2682	1.9269	1
Germany	0.1609	0.5418	1	0.1318	0.284	1	0.1346	0.312	1
Greece	0.0281	-0.9047	1	0.032	-0.9372	1	0.0318	-0.9311	1
Hungary	0.1105	-0.007	1	0.1134	0.059	1	0.1142	0.0649	1
Ireland	0.0877	-0.2556	1	0.1064	-0.0264	1	0.0972	-0.1407	1
Italy	0.0447	-0.7238	1	0.0397	-0.8424	1	0.0399	-0.833	1
Latvia	0.0797	-0.3428	1	0.1032	-0.0654	1	0.0998	-0.1086	1
Lithuania	0.0124	-1.0758	1	0.029	-0.9744	1	0.0274	-0.9839	1
Luxembourg	0.2383	1.3846	1	0.2181	1.3398	1	0.2254	1.41	1
Malta	0.0216	-0.9757	1	0.0215	-1.0655	1	0.023	-1.0368	1
Netherlands	0.1701	0.6419	1	0.149	0.494	1	0.1488	0.4837	1
Poland	0.0124	-1.0758	1	0.029	-0.9744	1	0.0274	-0.9839	1
Portugal	0.0263	-0.924	1	0.0246	-1.0283	1	0.0252	-1.0104	1
Romania	0.0071	-1.1339	1	0.0168	-1.1231	1	0.0163	-1.1179	1
Slovakia	0.0987	-0.1362	1	0.107	-0.0189	1	0.1087	-0.0011	1
Slovenia	0.1022	-0.0974	1	0.0953	-0.1621	1	0.0953	-0.1633	1
Spain	0.0343	-0.8369	1	0.0363	-0.8852	1	0.036	-0.8802	1
Sweden	0.1337	0.2448	1	0.1417	0.4048	1	0.1393	0.3686	1
United Kingdom	0.2727	1.7592	1	0.2562	1.8064	1	0.2498	1.7042	1

Figure 35: G-statistics and club memberships for the years 1995, 2008 and 2010.
Source: Own illustration in Excel according to the results obtained from Matlab.

4. Variables used in the Empirical Analysis

Variable	***Name***	***Details***	***Source***
GDP Growth	*GDPgrw*	Annual growth of Gross Domestic Product per capita.	Eurostat
GDP	*GDP*	Gross Domestic Product in per capita euro terms.	Eurostat
Unemployment	*Unemp*	Annual average unemployment as a % of total population.	Eurostat
Labor Force	*Labor*	Active population between 15 & 64 years as a percentage of total population.	Eurostat
Educational Attainment	*Edu*	Persons between 15 and 64 years, with upper secondary or tertiary education, as a percentage of the total population.	Eurostat
Government Expenditure	*Gov*	Total general government expenditure as a percentage of GDP.	Eurostat
Exports-Imports	*Eximp*	External balance of goods and services as a percentage of GDP.	Eurostat
Consumption	*Cons*	Final consumption expenditure in per capita euro terms.	Eurostat
R&D Expenditure	*R_D*	Total intramural R&D expenditure (GERD), for all sectors, as a percentage of GDP.	Eurostat

Figure 36: Set of variables used in the empirical analysis.
Source: Own illustration in Excel.

5. GDP graphs of the countries within the two clubs

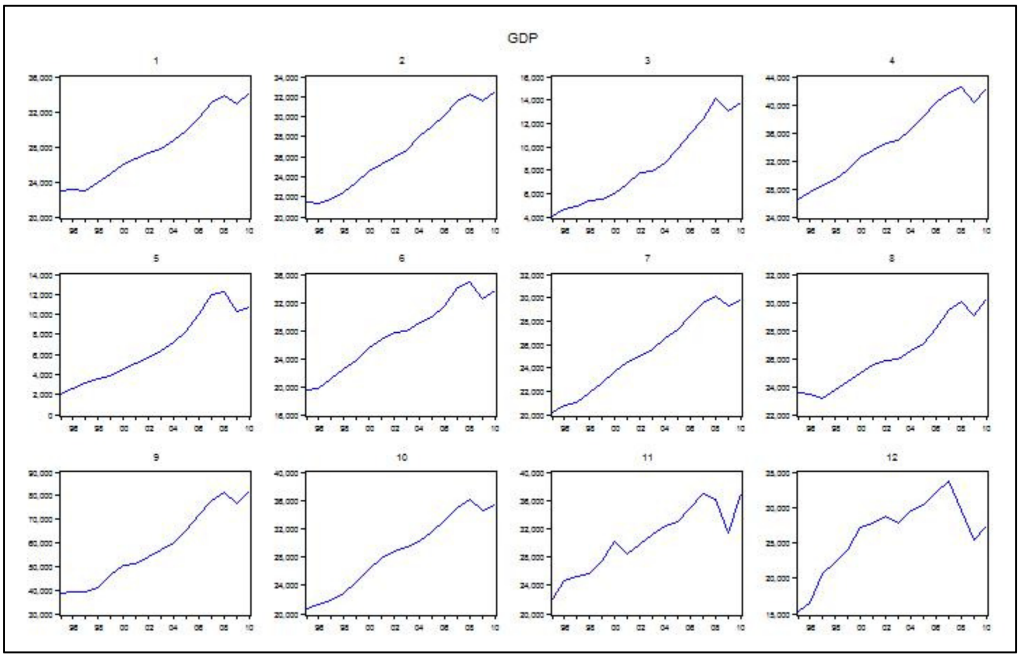

Figure 37: Evolution of Club 1 GDP per capita over time.
Source: EViews output.

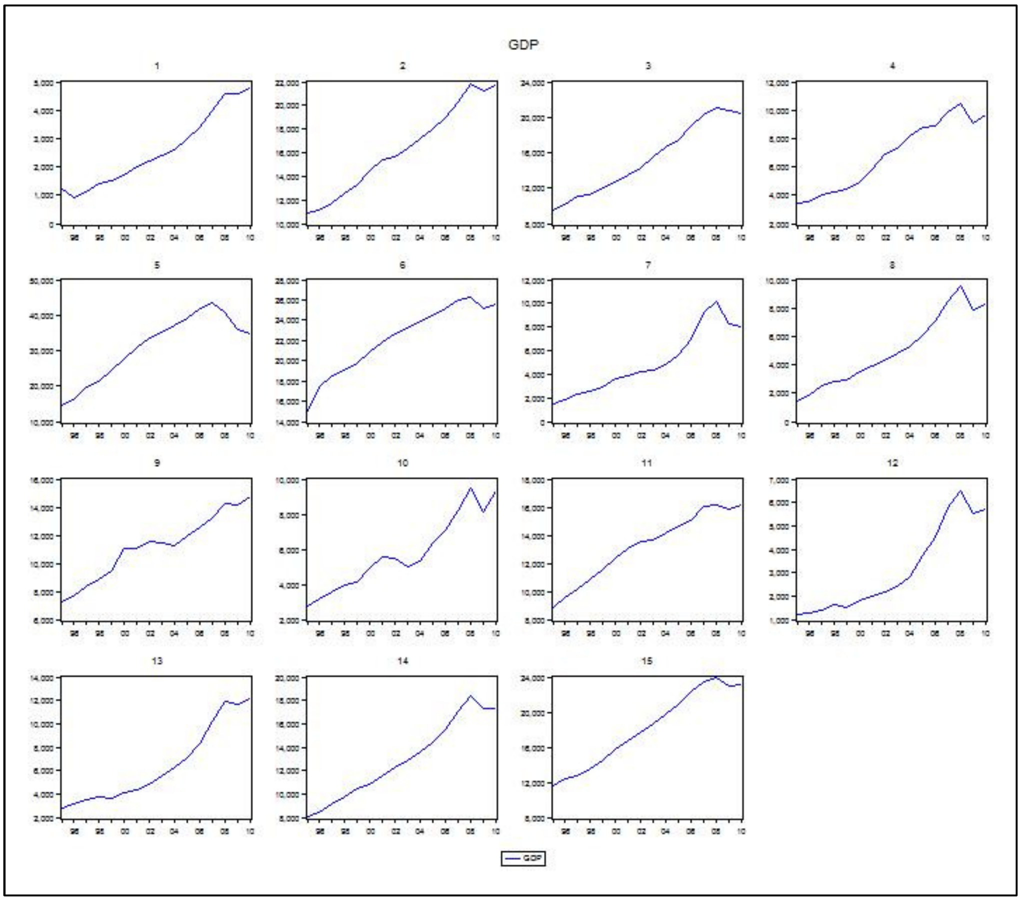

Figure 38: Evolution of Club 2 GDP per capita over time.
Source: EViews output.

6. Definitions

Measurement error: "… in this case we would like to measure the partial effect of a variable, but we can observe only an imperfect measure of it. "…when we plug this variable in the regression equation, we necessarily put a measurement error into the error term…" (Wooldridge, 2002).

Omitted Variables: "… appears when we would like to control for one or more additional variables but usually because of data unavailability, we cannot include them in a regression model…" (Wooldridge, 2002).

Simultaneity: "… arises when at least one of the explanatory variables is determined simultaneously along with the dependent variable…" (Wooldridge, 2002).

7. OLS Regression Analysis and the statistics

7.1. Performing regressions

For regression analysis purposes and to be able to test both the unconditional and conditional β-convergence hypothesis, the OLS (Ordinary Least Squares) estimation method is employed. Without going into much detail about the mathematical foundations of the OLS, it is important to mention certain specifics surrounding OLS estimation.

OLS is used for the estimation of linear regression models (such as the one shown in equation (13)) and aims at estimating the unknown parameters by minimizing the sum of the square vertical distances (ε_y) between the observed variables and the ones predicted by the linear model. The figure on the right illustrates these distances. However OLS is consistent in estimating β if and only if the error has a mean of zero and is uncorrelated with each of the explanatory variables. Otherwise the explanatory variables are said to be endogenous to the model, producing biased statistics. Endogeneity usually arises due to omitted variables, measurement errors and simultaneity (Wooldridge, 2002). Detailed definitions of these three biases are presented in Appendix 6.

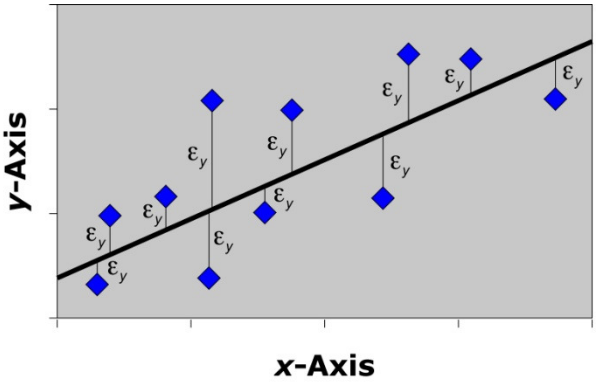

Figure 39: Error distances from the fitted regression line.
Source: Own illustration.

The choice of the OLS as a model for this regression analysis, relates to the linear nature of the assumed model, the effort to keep a certain level of simplicity in the analysis and to the fact that a relatively similar model had to be used for testing convergence in two quite different clubs. Nonetheless, empirical literature has shown that other estimation methods, such as Maximum Likelihood (ML), may perform better and produce more robust results. This is especially the case when testing at a regional level and when cases of substantive spatial dependence or spatial error dependence are tested (Fischer & Stirböck, 2006).

7.2. Fixed Effects modeling

The choice of using fixed effects modeling in regression analysis is important whenever there is an interest in analyzing the impact of variables that vary over time. Each entity (i.e. country) used in the regression analysis, has some specific characteristics that may or may not affect the predictor variable. For example, the political system of a country could affect the GDP level. As a result, the use of fixed effects is a way to control for those characteristics that may impact or bias the outcome of the analysis.

An important assumption of the fixed effects model is that since each entity is different, the individual characteristics (captured by the constant term) and the error term must not be correlated with the each other. If there is correlation then the fixed effects model is no longer suitable and maybe random effects of no effects models are more suitable. The Hausman test is used for such decisions between fixed and random effects modeling. A drawback of the fixed effects model is that it cannot be used to analyze the time-invariant causes of the dependent variable. In line with this assumption of the fixed effects modeling, countries' characteristics are perceived to be uncorrelated with their respective error terms.

To sum up, the fixed effect model controls for all those time-invariant differences between the entities (i.e. countries), in order to avoid for bias in the estimated coefficients due to omitted time-invariant characteristics. In this way the model can control for the impact of culture, religion, politics, etc. (Kohler & Frauke, 2009)

7.3. An insight into the statistics

This section contains a brief explanation of the statistics that are generated by EViews after running regressions.

Dependent Variable:
Method: Panel Least Squares

White cross-section standard errors & covariance (d.f. corrected)

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.555348	0.188639	2.943972	0.0038
C(1)	-0.038786	0.013729	-2.825167	0.0054
C(2)	-0.037048	0.014431	-2.567228	0.0113

Effects Specification: Cross-section fixed (dummy variables)

R-squared	0.918509	Mean dependent var	0.053879
Adjusted R-squared	0.907125	S.D. dependent var	0.053475
S.E. of regression	0.016297	Akaike info criterion	-5.276507
Sum squared resid	0.036119	Schwarz criterion	-4.885500
Log likelihood	431.5675	Hannan-Quinn criter.	-5.117697
F-statistic	80.67914	Durbin-Watson stat	2.132069
Prob(F-statistic)	0.000000		

Figure 40: Example of EViews' regression analysis output.
Source: Own illustration.

The first column contains the names of the variables used in the regression equation. The second column illustrates the estimated coefficients for each of these variables. The first coefficient is always the intercept of the equation and provides an estimation of α in equation (13). The second coefficient is the one describing the behavior of GDP per capita and represents β in the regression equation. The rest of the coefficients are estimates describing the behavior of the set of explanatory variables used in the model (the γ coefficients). The third column in EViews' output presents the estimated standard errors of the coefficients. These errors are a measure of the statistical reliability for the coefficient estimates. To be more precise, the larger the standard error, the more statistical noise is included in the coefficient estimates and the larger the probability to produce a biased model.

Column four and five of the statistical output are really important in the analysis. Column four presents an estimate of the t-statistic value. It is basically the ratio of the estimated coefficient to its standard error. The t-statistic is used to test the null hypothesis that the coefficient is equal to zero. However the t-statistic alone has no meaning, since it has to be examined together with the probability of observing it, given that the coefficient is equal to zero.

This probability is presented in column five. Assuming a 95% confidence interval for the statistical results of the regression, would mean that obtaining a t-statistic higher than 1.96 with a probability of less that 0.05 (i.e. 5%), makes that variable relevant and highly significant. Generally, the larger the t-statistic, the higher the relevance of the variable (provided that a small p-value is observed). In addition to the statistical output on the explanatory variables, EViews also outputs a set of summary statistics. Presented below is a brief description of each one of them (Quantitative Micro Software, LLC, 1994-2007).

R-squared: shows the amount of variance in Y (the dependent variable) that is explained by X (the set of independent variables) and basically measures the success of the regression equation in predicting the dependent variable. If the variables within the model have high explanatory power then the R^2 will be high and close to one. Otherwise the R^2 will be closer to zero.

Adjusted R-squared: has the same purpose as R^2, but penalizes for the addition of variables with no explanatory power to the model. Adjusted R^2 is never larger than R^2 and can decrease as more explanatory variables are added to the equation. For poorly fitting models, it can take negative values.

Standard Error of Regression: is a summary of the standard errors presented in column three and it is estimated based on the variance of the residuals.

F-statistic: tests the null hypothesis that the entire set of the slope coefficients in a regression (excluding the constant), are equal to zero. It has to be examined together with the probability of observing the F-statistic (**Prob-F**), which is the marginal significance level of the F-test. For the 95% confidence interval, if the probability is less than 0.05 then one can reject the above mentioned null hypothesis and the model is not biased. When comparing two models, if the F-statistic is larger with small probability value, then the model explain a lot more variability in the dependent variable, than the model with the smaller F-statistic.

Akaike Information Criterion (AIC): is a measure of the goodness of fit of the statistical model. It is said to describe the tradeoff between bias and variance in model construction, basically measuring how much information is lost in the effort of the model to describe reality. As such, the value of the Akaike criterion alone has no meaning. It has to be compared with an Akaike value generated from a different model in order to provide inferences. When comparing Akaike values, the smaller the value the better the model.

Shwarz Criterion: it is an alternative measure to the AIC with the difference of imposing a larger penalty for additional coefficients.

Durbin-Watson Statistic: is a measure used to detect serial autocorrelation in the residuals. The Durbin-Watson statistic takes values between 0 and 4. Values less than 1 and close to 0 is an indication of substantial positive serial correlation, meaning that the error terms are quite similar to each other. Values close to 4 are a strong indication of serial negative autocorrelation with the error terms differing significantly from one another. Values around 3 indicate no serial autocorrelation in the error terms. The presence of serial autocorrelation can affect the validity of a statistical model.

Bibliography

Anselin, L., & Rey, S. J. (1991). Properties of tests for spatial dependence in linear regression models. 23 (2), 112-131.

Arpaia, A., & Turrini, A. (2008). Government expenditure and economic growth in the EU: long-run tendencies and short-term adjustment. *Economic Papers 300* .

Barro, R. J. (1991). Economic growth in a cross-section of countries. *Quarterly Journal of Economics* , 106 (2), 407-443.

Barro, R. J., & Sala-i-Martin, X. (1992). Convergence. *Journal of Political Economy* , 100 (2), 223-251.

Barro, R. J., & Sala-i-Martin, X. (1991). Convergence Across states and Regions. *Brookings Papers on Economic Activity* , 22 (1), 107-182.

Blanchard, O. (2009). *Macroeconomics.* Upper Saddle River, New Jersey: Pearson Education International.

Canova, F. (2004). Testing for Convergence Clubs in Income per Capita: A Predictive Density Approach. *International Economic Review* , 45 (1), 49-77.

European Comission. (2011). *Inovation Union - Competitiveness Report 2011.* Brussels: European Comission.

Fischer, M. M., & Stirböck, C. (2006). Pan-European regional income growth and club-convergence. 40, 693-721.

Fischer, M. M., & Stirböck, C. (2006). Pan-European regional income growth and club-convergence. 40, 693-721.

Fischer, M. M., & Stumpner, P. (2008). Income Distribution dynamics and cross-region convergence in Europe. *Journal of Geographical Systems* , 10 (2), 109-139.

Getis, A. (2005). Spatial pattern analysis. *Encyclopedia of Social Measurement* , 3, 627-632.

Getis, A., & Ord, J. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis* , 24 (3), 189-206.

Jones, C. I. (2002). *Introduction to Economic Growth.* New York: W. W. Norton & Company Inc.

Kalyvitis, S. (No Date). *The Neoclassical Solow-Swan Growth Model.* Athens, Greece: Unpublished.

Kohler, U., & Frauke, K. (2009). *Data Analysis Using Stata, 2nd edition.* Stata Press.

Miller, R. (2008). Education and Economic Growth: From the 19th to the 21st Century. *Education and Economic Growth* , Cisco White Papers.

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

Quah, D. T. (1996). Empirics for Economic Growth and Convergence. *European Economic Review* , 40 (6), 1353-1375.

Quantitative Micro Software, LLC. (1994-2007). *EViews 6 User's Guide I & II.* USA, California: Quantitative Micro Software, LLC.

Richardson, H. (1978). *Regional and Urban Economics.* Harmondsworth: Penguin Books.

Romer, P. (1986). Increasing Returns and Long-run Growth. *Journal of Political Economy* , 94, 1002-1037.

Solow, R. M. (1956). A Contribution to the theory of Economic Growth. *Quarterly Journal of economics* , 70 (1), 65-94.

Tiiu, P., Kuusk, A., Schlitte, F., & Võrk, A. (2007). *Econometric Analysis of Income Convergence in selected EU countries and their NUTS 3 level Regions.* Tartu University Press.

Tselios, V. (2009). Growth and Convergence in Income per Capita and Income Inequality in Regions of the EU. *Spatial Economic Analysis* , 4 (3), 343-370.

Walterskirchen, E. (1999). *The Relationship between Growth, Employment and Unemployment in the EU.* Workshop in Barcelona.

Weil, D. N. (2009). *Economic Growth.* Boston: Pearson Addison Wesley.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data.* Cambridge, Massachusetts, London: The MIT Press.

Yanikkaya, H. (2003). Trade openness and economic growth: a cross-country empirical investigation. *Journal of Development Economics 72* , 57-89.

Yunlu, D., & Murphy, D. D. *R&D Intensity and Economic Recession: Investigating the moderating role of CEO characteristics.* Milwaukee.