Regression Tree Club Convergence Analysis

Nikolas Kuschnig, Lukas Vashold, Philipp Warum, Daniel Winkler

January 22, 2019

---

1 Introduction

One of the most discussed topics in macroeconomics, if not in economics overall, is the development of economies over time, be it on the national or on the regional level. Growth of these economies is usually associated with an increase of well-being of their respective inhabitants (e.g. Easterlin (1974), Stevenson and Wolfers (2008)) and is therefore a desirable objective. However, large disparities between the level of income per capita and the growth rates in the subsequent years of these economic entitites can be observed when comparing them to each other. In general, it is observable that regions with a lower initial level of income exhibit a higher rate of growth than richer regions. Several convergence theories emerged out of this observation, which try to address this phenomenon in different ways (e.g. Solow (1956), Swan (1956), Baumol (1986)). One of the most prevalent one is the *"club convergence hypothesis"* that pools economies which share similar initial conditions and structural characteristics in so-called convergence clubs. The regions in those clubs will converge to each other towards a common steady state of growth. This allows for multiple steady-states and implies convergence within these clubs but not necessarily between them.

There is a vast amount of empirical papers examining the club convergence hypothesis with European regions as the subject of study (e.g. Fischer and Stirböck (2006), Alexiadis (2012), Doran and Jordan (2013)) covering different timespans and identification methods that are used in order to pool regions into their corresponding club.

2 Testing for Convergence

2.1 Convergence Analysis

Analysing the economic development of regions of different sizes has a long tradition in the academic discourse between scholars and has often been the basis for policy decisions. Originating from the neoclassical theory with its general focus on equilibrating market forces, it has long been the dominant view that in the abscence of severe disruptions of these forces, there is a strong tendency towards income disparities to vanish, at least in the long-run. Therefore, it would only be natural for economies to converge towards each other (e.g. Solow (1956), Borts and Stein (1964)). This view has been challenged by the proponents of different schools of thought, which state that there is rather a divergence between economies if distribution is left to market forces alone. Because of the tendency of agglomeration of production factors and therefore also concentration of output, they argue that there will be a rather uneven development of regions than an even one (e.g. Myrdal and Sitohang (1957), Kaldor (1970)). Other important theories point out that times of convergence or divergence can alternate and that neither one of these processes is persistent over time but rather happen in episodes (e.g. Harvey (2006), Massey (1995)). Thus, following this logic, empirical studies should consider various timeframes and account for the different stages of the business cycle in order to get conclusive results. Furthermore, the convergence process seems to be dependent on the chosen level of observations (e.g. regional, national level). Convergence seems to be achievable on the national level, but, partly because of agglomeration and urbanisation effects within countries (Krugman 1991), divergence on the inter-regional level seems to be plausible (e.g. Smith and Timár (2010), Doran and Jordan (2013)).
With the revived interest in growth theories and the empirics behind those theories in the 1980s, two main concepts of measuring convergence (or divergence) have come up. The first one is the concept of β -convergence, which basically states that when regressing the regional growth rate of per capita income for a given time period on the level of per capita income of these regions in the first observation date, a negative coefficient would imply convergence. This stems from the logic that if the level of per capita income in the beginning of a certain time period has a negative effect on the growth rate in the subsequent years, regions (or countries) with a higher initial level would grow slower than the ones with lower income per capita and therefore they would converge towards each other. Vice versa, a positive coefficient would be a sign of divergence between the regions. The second concept, σ -convergence, is closely related to this notion, but focusses on the variance of levels of income per capita levels. Here, a set of regions is said to be converging if the variance of these levels is declining over a given time period. Thus, thanks to their relatedness one could say that *"[t]he existence of β -convergence will tend to generate declining dispersion or σ -convergence"* as is neatly summarised by Martin and Sunley (1998) among others. However, it should also be noted that for σ -convergence to take place β -convergence is necessary, but not sufficient, as the first one also depends on the error terms of the regressions conducted in order to quantify this effect. For an extensive discussion on the differences between these two concepts of concergence, see Furceri (2005). This thesis will adopt the notion of β -convergence as the concept that will be investigated in order to determine if there is evidence for income convergence among different clubs of European regions.

$$\frac{1}{T}\ln\frac{y_{it+T}}{y_{it}} = \alpha + \beta \ln y_{it} + u_i \tag{1}$$

Thus, equation (1) shows the basic setting for the regression used in the following analysis where y_t refers to income per capita at the beginning of the observation period, α can be interpreted as the steady state that the various clubs are converging to (if they do), u_i stands for the error term assumed to be centered around zero and to be normally distributed (this assumption will be reconsidered and altered in section **??**), β is the coefficient of interest and measures the speed of convergence. However, β cannot be interpreted right away, but has to be transformed in order to be interpretable. Following Barro and Martin (1992) this yields:

$$\beta^* = -\frac{1}{T} ln1 - T\beta \tag{2}$$

Estimating equation 2 with OLS and interpreting the coefficients in that way, thus *"averaging across [i] in a cross-section - regions, states, countries - constitutes the canonical β convergence analysis"* (Quah 1996). This kind of convergence analysis was conducted extensively with different regional foci and time periods (e.g. Barro and Martin (1995), Armstrong (1995) or Doran and Jordan (2013)). However, even though there was evidence for economic convergence between countries or regions that possess similar characteristics, this finding does in general not hold true when including countries or regions that do not share these characteristics.
Out of this recognition arose two hypotheses that were trying to deal with these implications. The *"conditional convergence hypothesis"* states that countries or regions that are similar in their structural characteristics even when they do not share the same initial conditions. Thus, every entity has their own, unique steady state that it converges to (Mankiw, Romer, and Weil 1992) depending for example on their respective rate of technology growth, institutional quality or population growth. Proponents of this approach argue that the neoclassical growth model brought forward by Solow (1956), Swan (1956) and similar naturally leads to the conditional convergence of income per capita as the most plausible hypothesis. However, as Galor (1996) argues on a theoretical level, the *"club convergence hypothesis"* is also a viable and competing hypothesis to conditional one. It differs from the first one in the regard that it allows for multiple steady-state equilibria and states that economies tend to converge when they exhibit similar structural characteristics and if they share similar initial conditions as well. In this sense it is needed to identify such clubs (described in more detail in subsection 2.2) and test for convergence in these clubs. Convergence between these clubs is no necessity, so disparities between them could persist or even extend. As one of the first studies to consider the possibility of different convergence clubs of regions, Baumol (1986) acknowledges that there is some evidence for this theory. Following studies like that of Quah (1996) support this view, at least to a certain extent. The empirics on club convergence were growing over time and newer contributions (e.g. Alexiadis (2012), Ghosh, Ghoshray, and Malki (2013)) also shed some new light onto these insights. While not every study comes to the conclusion that the convergence hypothesis is robust to certain aspects (for instance Fischer and Stirböck (2006) conclude that once controlling for spatial dependence, there seems to be no significant evidence for club convergence), different procedures were developed in order to identify such clubs effectively.

2.2 Club Identification

In order to identify convergence clubs, two approaches have been chosen. The first one focuses on an identfication based on differences in the initial conditions of a region as well as the initial conditions of regions that are surrounding this region. It can therefore be seen as a procedure that combines the regional distribution of initial conditions with the initial conditions of the region itself and is based on the G^*-statistic developed by Getis and Ord (1992). The second approach rather focuses on the transition paths of different regions and how they evolved over time. It tries to group regions depending on those transition paths and can therefore, in distinction to the first procedure, be seen as an approach that focuses rather on the individual evolvement of regions in a certain period without treating the spatial distribution of initial conditions as a key factor for its clubbing. The differences between these procedures makes it quite interesting to study both of them and compare their outcomes.