Regime Switching with Structural Breaks in Output Convergence

Fuat C. Beylunioğlu[†] Thanasis Stengos [‡] M. Ege Yazgan[§]

January 2, 2018

Abstract

In this paper, we examine empirically GDP per capita convergence using an approach that explicitly allows for regime switching in the long memory parameter d within the context of a Markov Switching (MS)–ARFIMA framework. As existing methods used in the estimation of standard MS models, such as the EM algorithm are no longer appropriate, we will make use of the Viterbi algorithm to estimate the long memory MS model used by Tsay and Härdle (2009). We will classify the output gap series into two regimes, a high d and a low d regime, where a high d close to unity would imply persistence and lack of convergence. By examining the path of d parameter over time which enables us to observe non-convergent behavior in more detail, we find that converging behavior is diminishing over time and divergence is the dominant force.

JEL Classification: C1, C2, O1, O4

Keywords: Markov Switching, long memory, structural breaks, output convergence, Viterbi Algorithm

[†]Istanbul Bilgi University, email: can.beylunioglu@bilgiedu.net

[‡]University of Guelph, email: tstengos@uoguelph.ca

[§]Istanbul Bilgi University, email: ege.yazgan@bilgi.edu.tr

1 Introduction

The concept of convergence (or lack of it) over time among different indicators of economic or social activity is central in the study of many economic phenomena. For example, one of the main predictions of (neoclassical) economic growth theory is that in the long run, all countries with similar technological characteristics would converge to a balanced growth path (steady state) equilibrium that will be entirely determined by the (exogenously) given growth rate of technical progress, which in turn would equal labour productivity growth. Hence economies with the same productivity would grow at the same rate and converge to the same equilibrium. This is the so called growth convergence hypothesis, which has been one of the main focal points of the empirical economic growth literature. In that context, a time series interpretation of the convergence hypothesis considers income gaps (or labour productivity gaps) between countries over time and analyzes whether these gaps would diminish, hence signifying convergence.

There have been a variety of models also known as "endogenous growth models" (in contrast to the traditional Solow model that relies on an "exogenous" technical change) that address directly such questions as why certain countries managed to grow faster than others, how the accumulation of human capital and R&D could enhance performance and why imperfect competition and international trade permit productivity gains that could not be achieved by closed economies with controlled markets among others, see Romer (1994) for an overview. Moreover, because of the many questions that have arisen theoretically and the advent of an abundant data base from a project of the World Bank (WB) there has been a huge increase of empirical activity in this area with many studies testing the validity and the predictions of the many models both endogenous and those of the augmented neoclassical with exogenous technical change growth variety. Finally, it has also been argued that the standard GDP index is unable by itself to capture the process of economic development and there may be other dimensions of well being such as life expectancy that also play an important role in explaining convergence among countries, see for example Stengos, Thompson, and Wu (2009) and Wu, Savvides, and Stengos (2014).

From a theoretical point of view, lack of convergence arises if there are constant or increasing returns to capital. In that case there may be a multiplicity of steady states (or absence of stable steady states) and a country's initial conditions will determine to which of these it will converge, see Azariadis and Drazen (1990) and Azariadis (1996). In essence, convergence to a single steady state implies that however poor, a country will inevitably converge to prosperity in the long run. In the absence of such a single steady state, poor countries may only converge to a common equilibrium with other poor countries and will never catch up with the prosperous ones.

The earlier literature on convergence has been surveyed by Durlauf, Johnson, and Temple (2005) and more recently by Johnson and Papageorgiou (2017). The time series approach to deal with the empirics of convergence was introduced by Bernard and Durlauf (1995, 1996) who cast it in terms of unit root and cointegration analysis.

Pesaran (2007) introduced a methodology based on a testing procedure that applies unit root tests to pairwise differences of the income per capita time series. Convergence is reached when the proportion of rejections obtained from the pairwise unit root tests is greater than a certain threshold. He applied this method to country groups belonging to different geographical regions and found no evidence of convergence clubs. Pesaran (2007) has extended the time series convergence concepts to the case where there is no requirement that the converging economies to be identical in all aspects including initial endowments. The main result is that for two economies to be convergent it is necessary that their output gap is stationary with a constant mean, irrespective of whether the individual country's output is trend stationary and/or contains unit root.

However, the above work assumes that the empirical analysis can be carried out within a $I(0)$ or $I(1)$ framework, yet it may be that a long memory framework is more appropriate. Michelacci and Zaffaroni (2000) introduce fractional integration but they restrict the long memory parameter to lie between (-0.5,0.5). Dufrénot, Mignon, and Naccache (2012) and more recently Stengos and Yazgan (2014b) use fractional integration analysis to test convergence for a group of developing countries. They employ an ARFIMA model and they allow the long-memory parameter d to be greater than 0.5. In other words, they do not simply restrict d to be in the interval (-0.5,0.5) but they also allow it to be between 0.5 and 1 as well as greater than 1. This gives rise to a rich classification of convergence cases and the authors are careful to examine the different cases that arise. Their analysis is contrasted with that of transient divergence, see Phillips and Sul (2007a,b), where convergence will take place eventually as divergent dynamics implied by idiosyncratic growth factors will diminish and will be dominated by the common components of economic growth. However, the analysis of Dufrénot et al. (2012) is subject to an important caveat, in that they fail to consider the issue of structural breaks that will affect the time series properties of the series under consideration. In the case of structural breaks, events that alter the steady state levels of per capita income will also change the mean reversion properties of relative outputs. This is the case of the work of Li and Papell (1999) for example. In the standard $I(0)/I(1)$ analysis, when structural breaks are present standard tests of convergence may lack power to reject the null of non-stationarity. The same will be true for an ARFIMA process where the presence of structural breaks may contaminate the dynamics and distort the estimation of d, the speed of convergence parameter. Stengos and Yazgan (2014a,b) extend the analysis of Dufrénot et al. (2012) to allow for structural breaks in the mean function of the gap series (gaps were measured as exchange rate differentials and output gaps respectively), but not in the speed of convergence parameter d. Furthermore, structural breaks were introduced in the mean function by a deterministic smooth mechanism that may not be appropriate as the regime switching forcing state variable may be latent and unobservable. Models that allow for different long memory regimes have been used in the literature but the regime switching is forced by an observable state variable, see Haldrup and Nielsen (2005).

Allowing for a latent unobservable regime forcing state variable can be accomplished in the context of a Markov Switching (MS) model, see Hamilton (1989). In a regime switching growth model, a growth process results from transitions between different regimes (states of nature) characterized by a common behavior of countries in a specific regime. Regime switching is country specific and takes place due to government policies or external shocks. The overall long-run growth rate of a country will depend on the time spent on each regime and countries with similar transition probabilities are grouped together. A simple application of an MS growth model by Kerekes (2012), is used to classify counties into a predetermined number of regimes (clusters) by transitioning between these different possible states within the context of a simple AR(1) process.

In the current paper we look at output gaps that may transition between different regimes (convergence and divergence) and we directly observe these transitions. Our framework of regime switching will attempt to unravel movements between convergent and divergent regimes for output gaps. Two countries converging or diverging over time will depend on the time that their output gaps have spent on each of these two regimes respectively. We will adopt a much richer MS framework than those adopted in the growth literature so far, allowing for a long memory framework. Following Tsay and Härdle (2009) we will make use of a Markov-Switching-ARFIMA (MS-ARFIMA) process which extends the hidden Markov model with a latent state variable, allowing for the different regimes to have different degrees of long memory. Recent papers in the literature have looked at changes in the persistence of a univariate time series, considering primarily a shift from a unit root process $I(1)$ to a stationary process $I(0)$ or vice versa at some unknown date over the sample under consideration. In that strand of the literature the analysis centers on the properties of estimators (and tests) in these extreme cases, see Perron et al. (2006) for a survey of testing procedures.

These models however deal with the extreme dichotomy of $I(1)$ versus $I(0)$ and do not allow for long memory and fractional integration. The main emphasis of the Tsay and Härdle (2009) approach has been to present a framework that can disentangle the impact of long memory dependence on the estimates of the latent regime parameters in the MS–ARFIMA framework. However, in their empirical application and simulations they did not consider regime switching in the long memory parameter but only in the mean and yet a long memory parameter regime switching may introduce contamination effects on the estimation of the mean parameters, see Diebold and Inoue (2001) for mentioning this concern. In a more recent paper Özkan, Stengos, and Yazgan (2016) using the Tsay and Härdle (2009) (MS-ARFIMA) framework of analysis conducted Monte Carlo simulations allowing for breaks both in the mean and long memory parameter. They found that breaks in the long memory parameter can have similar effects on the (in sample) fitting ability of the model irrespective of the presence of breaks in the mean parameter, confirming the contamination concern raised by Diebold and Inoue (2001). Hence, when considering the presence of structural breaks one should take into account their effect on the persistence parameter.

In this paper we propose an approach that allows for structural breaks within an MS-ARFIMA model and combines it with pairwise analysis of output gaps. Our contributions are as follows. Firstly, we will allow for structural breaks in the persistence parameter following the evidence of Özkan et al. (2016). We make use of the Viterbi algorithm of Tsay and Härdle (2009) to estimate the long memory MS model, as existing methods used in the estimation of standard MS models, such as the EM algorithm used by Hamilton (1989), are no longer appropriate. Secondly, by examining the path of the d parameter over time enables us to observe directly the movement between the two regimes (especially the transition from convergence to divergence) for given pairs. Finally, we will classify the output gap series into two regimes, a high d and a low d regime, where a high d of unity or above unity would imply persistence and lack of convergence. The convergence (or lack of it) for the group of countries under consideration within the pair-wise testing framework that we consider will be determined by the overall proportion of pairs showing stationary behaviors.

Even though our approach does not involve explicit testing of particular hypotheses regarding the order of integration, our results are similar to the findings of Pesaran (2007) and Stengos and Yazgan (2014b), where no significant convergence is detected. In fact by examining the path of d parameter over time, something that enables us to observe non-convergent behavior directly, we find that converging behavior is diminishing over time and divergence is the dominant force. Our findings are in agreement with the broad literature on the subject where with the exception of some early studies that have been criticized extensively due to econometric problems, there is a broad consensus for the lack of convergence, see Johnson and Papageorgiou (2017). As it is argued there the process of convergence is not smooth but rather characterized by ”start and stop” behavior. It is then possible that several mechanisms of divergence and convergence are concurrently at work across countries in different stages of their development process and plotting the evolution of the long memory parameter allows us to directly observe such behavior.

The paper is organized as follows. In the next section we present the empirical methodology that we follow as well as an explicit presentation of the algorithm that is used to estimate the MS-ARFIMA model with the Viterbi algorithm. In the following section we will present our empirical findings and then we will conclude.

2 Methodology

2.1 Pairwise Method

Suppose that the log GDP per capita series of country i and j at time t are as follows

$$Z_t^{ij} = y_t^i - y_t^j = \mu(t) + \varepsilon_t \quad \sim I(d),\ i = 1, ..., N-1,\ j = i+1, ..., N,\ t = 1, ..., T, \quad (1)$$

where T is the length of time interval, N is the number of countries and y_t^i and y_t^j denotes the log GDP per capita series of i and j. ε_t stands for the disturbance term and d is the long memory parameter. Here $\mu(t)$ can represent a constant or a function of time as well. (see Stengos and Yazgan (2014b)). In the simple $I(0)/I(1)$ framework the two log GDP per capita series will be drifting together overtime if $\varepsilon_t \sim I(0)$ and it is appropriate to assert that countries i and j are convergent. On the other hand, if $\varepsilon_t \sim I(1)$, a nonstationary process would indicate that the log difference series between i and j is nonstationary and the two log GDP per capita series would be drifting apart over time, indicating that countries i and j are not converging.

However, when there are more than two countries, there is uncertainty in determining whether countries are converging altogether to a steady state. In the literature, the main approach centers on testing if all countries in the group are converging to the group average or a chosen country as a benchmark (generally United States), hence applying unit root tests to the pairwise differences of each group member with the average or the selected benchmark country. Alternatively, another approach is to apply multivariate stationarity tests to determine convergence. The former approach is criticized for the arbitrariness in choosing the benchmark country or the country average, while the latter is not preferred because of the difficulties in applying it to large groups.

To offer a possible remedy to both of the above difficulties Pesaran (2007) suggested to use all pairs $N(N-1)/2$ instead of a benchmark or an average. According to this approach, if one tests for convergence of a group of N countries, all

$N(N-1)/2$ pairs are subjected to unit root testing. Pesaran (2007) showed that, if a group of N countries are non-convergent, the rejection rate of the null hypothesis of non-stationarity calculated by $N(N-1)/2$ tests is equal to the nominal size of the individual tests, i.e. the probability of Type 1 error[1]. However, this approach is also subject to some problems. The nominal size of the tests may differ from its actual level and there may be distortions as for the rejection rate to converge to α in the limit under the null hypothesis requires for if N and T to tend to infinity. That of course would not be the case if they are relatively small in a given application.

Going beyond the $I(0)/I(1)$ framework, in the case with $I(d)$, where d now refers to a fractional integration parameter $-0.5 < d \leq 1$, Dufrénot et al. (2012) and Stengos and Yazgan (2014b) suggested a much richer classification on which one can distinguish between the different convergence cases. Different values of d define different types of convergence and we enumerate these different convergence cases below.

Case 1: $-0.5 < d \leq 0$. This is the case of a short memory process, where there is "fast catching-up" or "short memory catching-up".

Case 2: $0 < d < 0.5$. This is the case of a long memory process, but still stationary process, where there is a slow or smooth decay in the catching-up process. Here, output differences in the remote past will linger on in the current output difference, although with a smaller influence. This is the situation when a country spends a long time on a transition path towards a common long-run trend.

Case 3: $0.5 < d < 1$. This is the case of a long memory process, which is non-stationary but still mean reverting. In that case the process is characterized by high persistence, whereby any output differences in the distant past will still have a long-lasting influence in the present.

Case 4: $d \geq 1$. This is the case of a unit root or an explosive process, where any initial difference is not expected to be reversed in the future or there is a strong magnification effect.

Stengos and Yazgan (2014b) and Dufrénot et al. (2012) examine convergence based on estimating d using a number of different estimators and then testing sequentially in which of the above cases each output gap pair[2] belongs to.

---

[1] More specifically, it is shown that under the null hypothesis of N countries being non-convergent, the rejection rate of individual tests converges to the nominal size, α, as N and $T \to \infty$, even though individual tests are not independent cross- sectionally. Thus, in order to reject non-convergence of N countries, it is enough to show that the proportion of rejections over $N(N-1)/2$ tests is larger than the significance level of individual tests. In that case for example, if the significance level is 5%, the proportion of rejections must exceed 0.05.

[2] Defined as by using all pairs as in Pesaran (2007) approach (Stengos and Yazgan (2014b)) or by using a benchmark country (Dufrénot et al. (2012)).

Stengos and Yazgan (2014a,b) allow for structural breaks in the mean function of the gap series (gaps were measured as exchange rate differentials and output gaps respectively), but not in the speed of convergence parameter d. Furthermore, structural breaks were introduced in the mean function by a deterministic smooth mechanism that may not be appropriate as the regime switching forcing state variable may be latent and unobservable.

As in the case of the size distortions that may affect the Pesaran (2007) testing approach, similar concerns are present in the approach of Dufrénot et al. (2012) and Stengos and Yazgan (2014b), not to mention the additional issues of low power and size distortions that are present in any sequential testing methodology. To avoid the above potentially serious problems that may affect the sequential testing approach outlined above we proceed to simply classify the estimates of the persistence parameter d of the different pairs of output gaps in the presence of regime switching as being consistent with non-converging or converging behavior. In that sense, within the context of an MS–ARFIMA framework and will proceed to analyze the issue of convergence by identifying a long memory regime with $d < 1$ and a nonstationary unit root or explosive regime where $d \geq 1$ in the movements in the pairwise gap series allowing for d to move between regimes. Hence, examining regime switching based on d will allow us to distinguish between high persistent regimes as those that are consistent with unit root behavior[3]. Moreover the MS approach will enable the long memory parameter to be traced back in time and provide more information about convergence or factors that affect it. We will use a framework that relies on pair-wise analysis of all possible pairs with all possible estimates of d, without the need of a benchmark (country). For each time point, the long-memory parameter estimates of the output gap series for all available pairs are grouped with respect to their position against the cut off level $d' = 1$ and the presence of convergence is determined from the overall results of these groups.

The presence of convergence is determined from the overall results by looking at all the $N(N-1)/2$ pairs of the N countries that we have. In fact by examining the path of d parameter over time, something that enables us to observe non-convergent behavior directly we find that converging behavior is diminishing over time and divergence is the dominant force. Our approach is based on the estimation of d and its (potential) movement between regimes. In that sense we treat the d estimates in each regime as separate entities as we have not tested for the significance of the break as the (asymptotic) properties of such a test would be unknown. More

[3]The main aim of this study is to introduce a new methodology that captivates regime switches in output gaps. Even though an analysis considering more than two states may be useful to differentiate between stationary, mean reverting and nonstationary processes which in return allow for more precise clustering, we will leave that for future research plans.

importantly, we avoid the potentially severe size distortions that would affect the results of any approach based on testing as argued above.

In the next section we present the Virerbi algorithm that forms the basis for the estimation method of the MS model that allows for regime switching in the long memory parameter d and its estimation. The description of the method is presented in order to clarify its usefulness as it applies to our model and it differs from how it was presented in Tsay and Härdle (2009) who also use Viterbi (1967).

2.2 The Viterbi Algorithm

In order to allow the gap to be state switching, we will consider each Z_t^{ij} in Equation 1 following an MS-ARFIMA$(0,d,0)$ as

$$Z_t = \mu + (1-L)^{-d_{s_t}}\,\varepsilon_t, \tag{2}$$

where $\varepsilon_t \sim N(0,\sigma)$ $I.I.D$, $L(\cdot)$ is the lagging operator and d_{st} is long memory parameter. Notice that we omit i and j in order to simplify the notation and μ and σ are constant for each pairs, allowing only d_{s_t} to vary over time and among states.

Considering T observations, we assume that gap series is $\{Z_t\}_{t=1}^{T}$ is driven by a filtration process $\{s_t\}_{t=1}^{T}$ where $s_t \in \{H,L\}$ indicates state of regime at time t, either high and low respectively. We also assume that the states are following a Markov property according to the transition matrix,

$$\mathscr{P} = \begin{bmatrix} P_{HH} & P_{HL} \\ P_{LH} & P_{LL} \end{bmatrix},$$

where $P_{ij} = \mathbb{P}(s_t = j \,|\, s_{t-1} = i)$ and $P_{iH} + P_{iL} = 1$ for $i = H,L$. Hence it is implied that the transition probabilities are constant over time and there is no memory;

$$\mathbb{P}(s_t|s_{t-1}, s_{t-2}, \ldots, s_1) = \mathbb{P}(s_t|s_{t-1}),$$

that is each state only depends on the previous realization regardless of the history of the states.

With above characteristics, observation set $Z_{1:T} = \{Z_1, Z_2, \ldots, Z_T\}$ is an output of $s_{1:T} = \{s_1, s_2, \ldots, s_T\}$ and a constant parameter vector $\zeta = \{\mu, \sigma, d_H, d_L, \mathbb{P}_{HH}, \mathbb{P}_{LL}\}$. However in reality we do not observe the state series $s_{1:T} = \{s_1, s_2, \ldots, s_T\}$ but only the outputs $Z_{1:T} = \{Z_1, Z_2, \ldots, Z_T\}$ and should estimate the most likely path of states hidden behind observations. The estimation is not easy since there is one most likely path among 2^T possibilities; more clearly it is computationally difficult to evaluate all paths to determine the most likely one. To this end, we will adopt MS-ARFIMA(p,d,q) model proposed by Tsay and Härdle

(2009) which makes use of Viterbi algorithm to estimate Hidden Markov states. Below, we will follow Tsay and Härdle (2009) and Viterbi (1967) to construct the model.

Instead of evaluating all possible paths, Viterbi algorithm calculates probabilities for each step recursively and determines the most likely path passing through each node. More clearly, since we have two regimes at each step, we have two best paths (survivors) that ends up with H and L respectively. To exemplify, for $t = 2$, the algorithm evaluates all four paths conditioned to results of previous step and chooses two survivors; thus for the next step, $t = 3$, we end up with four possible paths with two prospective survivors. This evaluation continues to the last step where one best route having the greater probability is chosen. To express our problem mathematically, we first define the objective by

$$\max_{s_{1:T}} \mathbb{P}(s_{1:T}|Z_{1:T}) = \max_{s_{1:T}} \mathbb{P}(s_{1:T}, Z_{1:T}), \tag{3}$$

and the maximizer

$$s^*_{1:T} = \{s^*_1, s^*_2, \ldots, s^*_T\} = \arg\max_{s_{1:T}} \mathbb{P}(s_{1:T}, Z_{1:T}), \tag{4}$$

as the most likely path. To find the maximum value of $\mathbb{P}(s_{1:T}, Z_{1:T})$ and the maximizer $s^*_{1:T}$, we define a recursive function $M_t(s_t) = \max_{s_{1:t-1}} \mathbb{P}(Z_{1:t}, s_{1:t})$ for $t \in (1, T]$ and rewrite $M_t(s_t)$ as

$$\begin{aligned} M_t(s_t) &= \max_{s_{1:t-1}} \mathbb{P}(Z_{1:t}|s_{1:t})\mathbb{P}(s_{1:t}|s_{1:t-1}, Z_{1:t-1})\mathbb{P}(s_{1:t-1}, Z_{1:t-1}), \\ &= \max_{s_{t-1}} \mathbb{P}(Z_t|s_t)\mathbb{P}(s_t|s_{t-1}) \max_{s_{1:t-2}} \mathbb{P}(s_{1:t-1}, Z_{1:t-1}), \\ &= \max_{s_{t-1}} \mathbb{P}(Z_t|s_t)\mathbb{P}(s_t|s_{t-1})M_{t-1}(s_{t-1}). \end{aligned} \tag{5}$$

Notice that in the above equation, second component is given by the transition probability matrix and constant over time; and the recursion is provided by the last component that depends on the previous outcome. On the boundary $t = 1$, however, we have

$$M_1(s_1) = \mathbb{P}(Z_1|s_1)\mathbb{P}(s_1), \tag{6}$$

where $\mathbb{P}(s_1)$ is chosen as the limiting probability obtained via transition matrix. Thus, at each step we evaluate the paths passing through nodes and choose two paths for each $s_t \in \{H, L\}$ by $s^*_{1:t-1} = \arg M(s_t)$. Once we have complete the evaluation step by step, we end up with two filtration processes with their likelihood estimations among which we choose the one having higher likelihood estimation.

In (5), only term that links observations and states is $\mathbb{P}(Z_t|s_t)$. In our case this probability can directly be estimated via (2) if ζ is given. To this end, joining

Durbin-Levinson recursion and rearranging the likelihood function for a given path gives,

$$L(s_{1:T}, Z_{1:t}; \zeta) = \prod_{t=1}^{T} (2\pi)^{-1/2} v_{t-1}^{-1/2} \exp\left\{ -\frac{(Z_t - \hat{Z}_t)^2}{2v_{t-1}} \right\} \mathbb{P}(s_t | s_{t-1}), \tag{7}$$

where $\hat{Z}_t$ is the one step prediction of Z_t and v_{t-1} is the corresponding variance prediction. Recursive structure of this function enables the likelihood estimation to be integrated into Viterbi algorithm derived earlier. This combination can be done simply by taking,

$$\mathbb{P}(Z_t | s_t) = (2\pi)^{-1/2} v_{t-1}^{-1/2}(s_t) \exp\left\{ -\frac{(Z_t - \hat{Z}_t(s_t))^2}{2v_{t-1}(s_t)} \right\}, \tag{8}$$

where the predictions $\hat{Z}_t(s_t)$ and $v_{t-1}(s_t)$ are functions of the state variable. Thus, for a given parameter vector ζ, the above procedure yields the best path $s_{1:T}^*$ having the largest likelihood. Lastly, the estimation of ζ amongst infinite set of parameters can be performed numerically by using an optimization software. We made use of the R programming language with its optim function installed in base library[4].

The properties of the above algorithm within the context of an MS-ARFIMA model were with structural breaks both in the mean and the long memory parameter were examined recently by Özkan et al. (2016) by means of an extensive Monte Carlo simulation. The estimates of both the mean function and the long memory parameter were found to be well behaved in the presence of structural breaks and the Viterbi algorithm performed very well in terms of stability and convergence.

3 Data and Empirical Findings

We update the Maddison data set that was used in Stengos and Yazgan (2014b). The Maddison data consist of annual GDP per capita data for different time periods and for different country groups depending on the period. The period covering 1950 to 2010 for example consists of 141 countries[5]. The country coverage is given in

[4]The computations were made with a 2 cores/4 threads i76500U Windows 10 64bit computer and we used mclapply.hack function (an adaptation of mclapply in R for Linux) to allow multiprocess computing in Windows. For the 1930 data with 630 pairs with 81 time-points each, it took 4520 seconds (∼72 minutes). However note that the required time reduces slightly with higher core/thread processors such as regular i7 processors with 4 cores.

[5]The data come from the Maddison Project, 2013 version. Some countries are missing observations at the end of the period in the final two years. The data are available at http://www.ggdc.net/maddison/maddison-project/home.htm, and they include all possible countries available.

Table 1 for all groups considered[6]. We considered seven different data sets, three based on data availability starting from 1930, 1940 and 1950. The other three are subsets of 1950 where countries are clustered according to membership in Europe, the Group of Seven (G7) and inclusion of S&P Emerging Markets classification[7].

In order to illustrate the usefulness of our approach in following the movements between the low persistence and the high persistence regime states over time we also report the plots of d for certain pairs of countries, both developed and developing countries with different patterns of bilateral convergence behavior. These special cases of pairs can be found in all the data sets that we used and they are representative of the patterns that we obtain tracing the history of their respective d estimates over time. Using $d = 1$ as the threshold that acts as a demarcation between the low persistence regime with the absence of a unit root (C) and the high persistence regime with the presence of a unit root (D) we can trace the over time history of regime switching. Figure-1, Figure-2 Figure-3 plot a sample of cases for pairs of European country pairs, developed and developing country pairs respectively. Whereas for the European group there is evidence of convergence, this is not the case for the group of developing countries in these plots. For example pairs such as Peru and the Philippines display episodes of convergence in the 1980's and then they diverge, while Peru and South Africa converge until 1980 and then after a short oscillation between divergence and convergence in the 1980's they diverge after that[8]. The figures for developed countries suggest how such pairs as Belgium and Netherlands or Sweden and the US have had an estimate of d of their output gap become large during WWII and then drop to convergent behavior values for the after the WWII period. For another pair, the US and Canada there is an oscillation that captures the 1930's and WWII. This is an interesting way of identifying very significant events that have influenced regime switching in world history.

We proceed to consider the transition between states as follows. Table 2 displays the summary of results for the different datasets. The outcomes are displayed in five groups indicating the different states and the transition between states that are considered in our analysis. As before, we define two states, low (L) and high (H). The low state occurs when there is a low value of d consistent with the absence of a unit root (*C)* and consequently consistent with a type of convergence for the given pair of countries, whereas the high state occurs when there is a unit

[6]Notice that "1930" and "1940" groups are identical in terms of the countries included, however, naturally differ in terms of data length.

[7]The reason that we have looked at the latter groups is to ensure a degree of homogeneity within each group of countries in order to give the convergence process its best chance.

[8]For all the plots that are presented in the above figures we carried out likelihood ratio tests for the null hypothesis of no breaks. All the test statistics strongly rejected the null hypothesis with p-values of zero.

root in the output difference (*D)* and consequently divergence. More precisely, the cases displayed with a single letter in the second column indicate stable regimes, i.e. $d_L < 1$ for (C) and $d_H \geq 1$ for (D). The other cases denote the transition states between C or D. In that case, $C-D$ stands for the cases of transitions from the low to a high state, in other words from converging to diverging behavior, moving from d estimates such that $d_L < 1$ to $d_H \geq 1$. The case $C-C$ denotes the transition from converging to converging behavior, while $D-D$ the case of for diverging to diverging behavior, with $d_L, d_L < 1$ and $d_H, d_H \geq 1$ respectively[9]. The third column displays distribution of percentage over all $N(N-1)/2$ pairs resulted in each group, and the following columns displays average parameter estimates.

The results are similar to the findings of Pesaran (2007) and Stengos and Yazgan (2014b), where no significant convergence is detected. In Table 2, the proportion of pairs with $d < 1$ are around 10% for the 1930 data set, but it is much lower for the 1950 data. Furthermore, for each dataset a large proportion of pairs yield d estimates that are greater than 1 and this proportion gets even larger for the 1950 group of countries. The usefulness of the MS-ARFIMA estimation framework becomes clearer in the case of the $(C-D)$ transition results, that is from a converging to diverging behavior for countries in this group. For example, the estimates for the 1930 data show that for 11.4% of pairs (9.2% of the C group plus 2.2% of the C-C group) have $d < 1$, whereas for 52.4% (43.2% of D group and 9.2% of the D-D group) $d \geq 1$. For the same data set 36.2% of pairs belong to the $C-D$ group with the averages $d_L = 0.465$ and $d_H = 1.281$ respectively. Clearly in that case, there is a movement from what would appear otherwise a convergence behavior with a low estimate of d, to a divergent one. It is worth noting that the convergent single state as well as the transition of a convergent state to another convergent state diminishes as expected as we move from more homogeneous to less homogeneous groups. In general we observe that for the larger 1950 group of countries, there are lower percentages for C and $C-C$ cases which implies more incidents of non-convergence in these data set compared with the earlier 1930 and 1940 data. As mentioned earlier, the analysis centers on the use of $d = 1$ as this threshold creates the low persistence (C) and high persistence (D) regimes. This cut off point is of course arbitrary and estimates of d that are very close to unity from below for all intensive purposes will be similar to the ones that are slightly above unity. However, as it is found in our results there is a tendency to move from (C) to (D) and as such it is not the individual estimates of d that matter but the overall tendency for the transitions form lower

[9] Our approach is based on the estimation of d and its (potential) movement between regimes. In that sense we treat the d estimates in each regime as separate entities as we have not tested for the significance of the break at $d_L = 1$, since the (asymptotic) properties of the d estimates are not known. The approach that we follow simply centers on the value of d being unity and we ignore the fact that d may bifurcate between regimes as we treat the two values as separate estimates.

to higher values that emerges as our main finding, supporting divergent behavior.

The MS-ARFIMA model using the Viterbi algorithm also produces the path of d parameter over time which enables us to observe non-convergent behavior in more detail. For Figure-7 we first obtained d series for each country pair and calculated the cross-sectional proportion of $d < 1$ for each time point t. In the figure it is seen that the proportion for which $d < 1$ is diminishing over time, an indication that divergence is a entropy type phenomenon over time, with divergence being the dominant force.

4 Robustness

The empirical findings of the previous section are based on the premise that μ was constant for all pairs allowing only d_{s_t} to vary over time and among states. As mentioned in the introduction, even though regime switching in the long memory parameter may introduce contamination effects on the estimation of the mean parameters, ignoring the opposite effect of breaks in the mean parameter may also introduce contamination on the estimation of the long memory parameters. In order to control the presence of such a case in this section we repeat the analysis of the previous section by allowing breaks in both μ and d_{s_t}.

Table 3 displays the summary of results for the different datasets. The outcomes are again displayed in five groups indicating the different states and the transition between states that are considered in our analysis. As before, we define two states, low (*L*) and high (*H*). The low state occurs when there is a low value of d consistent with the absence of a unit root (*C)* and consequently consistent with a type of convergence for the given pair of countries, whereas the high state occurs when there is a unit root in the output difference (*D)* and consequently divergence. The results are very similar to the ones found in Table 2 with all the categories displaying similar patterns of overall divergence. Europe as a separate group displays the highest proportion of convergent pairs, 7.4%, yet, the overwhelming majority of pairs still displayed divergent behavior. In order to check further for the robustness of our previous findings we also produced the same figures plotting the movements of d between the C and D states over time for the same pairs of countries, both developed and developing with different patterns of bilateral convergence behavior that were analyzed earlier, for pairs of European, developed and developing countries, see Figure-5, Figure-4 and Figure-6 respectively. Figure-8 displays the evolution of proportion of convergent pairs through time similar to Figure-7. The patterns and the plots obtained in the case where we allow breaks in both μ and d_{s_t} are very similar to the ones obtained before when μ was constant.

5 Conclusion

In this paper, we examined empirically GDP per capita convergence using an approach that explicitly allows for regime switching in the long memory parameter d within the context of a Markov Switching (MS)–ARFIMA framework combined with testing of the unit root hypothesis using ADF tests of pair-wise output gaps. Following Tsay and Härdle (2009) we will make use of the Viterbi algorithm to estimate this model and we classify the output gap series into two regimes, a high d and a low d regime, where a high d close to unity would imply persistence and lack of convergence. Our results are similar to the findings of Pesaran (2007) and Stengos and Yazgan (2014b), where no significant convergence is detected. In fact by examining the path of d parameter over time, something that enables us to observe non-convergent behavior in more detail, we find that converging behavior is diminishing over time and divergence is the dominant force. Our findings are in agreement with the conclusions of Johnson and Papageorgiou (2017) who have produced the most comprehensive survey on the subject of convergence up to date who argued that several mechanisms of divergence and convergence may be concurrently at work across countries in different stages of their development process. From a methodological point of view our method enables us to plot the evolution of the long memory parameter over time and allows us to directly observe such behavior.

Tables and Figures

Table 1: Country Groups based on Economic Characteristics and Data Availability for Growth Application

1930 & 1940	Germany, USA, Argentina, Australia, Austria, Belgium, UK, Brazil, Denmark, Ecuador, Finland, France, Guatemala, South Africa, India, Netherlands, Ireland, Spain, Sweden, Switzerland, Italy, Japan, Canada, Colombia, Costa Rica, Mexico, Norway, Peru, Portugal, Sri Lanka, Chile, Turkey, Uruguay, Venezuela, New Zealand, Greece
1950	1930 & 1940 + Albania, Bulgaria, Czecho-Slovakia, Hungary, Poland, Romania, Bolivia, Costa Rica, Dominican Rep., Jamaica, China, Indonesia, Philippines, South.Korea, Thailand, Taiwan, Bangladesh, Burma, Hong.Kong, Malaysia, Pakistan, Singapore, Sri Lanka, Cambodia, Vietnam, Bahrain, Iran, Iraq, Israel, Jordan, Kuwait, Oman, Qatar, Saudi.Arabia, Syria, UAE, Yemen, Algeria, Angola, Burkina Faso, Cameroon, Ivory Coast, Egypt, Ethiopia, Ghana, Kenya, Madagascar, Malawi, Mali, Morocco, Mozambique, Niger, Nigeria, Senegal, South Africa, Sudan, Tanzania, Tunisia, Uganda, Congo.Kinshasa, Zambia, Zimbabwe
Europe	Austria, Belgium, Denmark, Finland, France, Germany, Italy, Netherlands, Norway, Sweden, Switzerland, UK, Ireland, Greece, Portugal, Spain, Albania, Bulgaria, Hungary, Poland, Romania
G7	Canada, France, Germany, Italy, Japan, UK, USA
S&P	Brazil, Chile, Colombia, Mexico, Peru, Hungary, Poland, China, India, Philippines, Thailand, Taiwan, Malaysia, Turkey, Egypt, Fas, South Africa

Table 2: MS-ARFIMA$(0,d,0)$ breaks in d only results

Data (T,N)	Conv.	Perc.	Avg. d_1	Avg. d_2	Avg. P_{11}	Avg. P_{22}	Avg. μ	Avg. σ
	C	9.2%	0.917	-	0.998	-	0.887	0.067
	D	43.2%	1.199	-	0.999	-	0.380	0.067
1930	C - C	2.2%	0.944	0.096	0.990	0.874	0.742	0.058
(81,36)	C - D	36.2%	1.281	0.465	0.981	0.882	0.378	0.057
	D - D	9.2%	1.531	1.200	0.886	0.946	0.514	0.048
	C	8.1%	0.907	-	0.998	-	0.776	0.070
	D	54.1%	1.214	-	0.999	-	0.413	0.058
1940	C - C	2.5%	0.958	0.113	0.987	0.853	0.733	0.058
(71,36)	C - D	26.3%	1.385	0.151	0.980	0.907	0.243	0.055
	D - D	8.9%	1.681	1.232	0.906	0.967	0.159	0.048
	C	7.7%	0.894	-	0.997	-	0.764	0.068
	D	68.7%	1.259	-	0.999	-	0.567	0.058
1950	C - C	1.3%	0.925	0.485	0.983	0.844	1.453	0.067
(61,95)	C - D	18.3%	1.246	0.356	0.985	0.871	0.665	0.065
	D - D	4.1%	1.325	1.136	0.976	0.871	0.726	0.053
	C	5.6%	0.904	-	0.999	-	0.212	0.044
	D	77.5%	1.249	-	0.999	-	0.146	0.034
Europe	C - C	0.4%	0.956	0.754	0.982	0.897	1.208	0.039
(61,22)	C - D	10.8%	1.298	0.523	0.982	0.879	-0.155	0.030
	D - D	5.6%	1.363	1.170	0.971	0.890	-0.681	0.033
	C	6.3%	0.916	-	0.999	-	0.322	0.044
	D	78.7%	1.260	-	0.999	-	0.225	0.034
Europe+G7	C - C	0.7%	0.946	0.752	0.991	0.875	1.371	0.042
(61,25)	C - D	10.3%	1.256	0.511	0.982	0.878	0.165	0.029
	D - D	4.0%	1.382	1.193	0.971	0.883	-0.359	0.032
	C	3.9%	0.912	-	0.999	-	-1.216	0.048
	D	80.3%	1.278	-	0.999	-	0.496	0.041
Europe+S&P	C - C	0.2%	0.956	0.754	0.982	0.897	1.208	0.039
(61,37)	C - D	11.1%	1.301	0.514	0.980	0.885	0.391	0.039
	D - D	4.5%	1.359	1.158	0.976	0.878	0.177	0.038
	C	4.9%	0.944	-	0.998	-	-2.383	0.048
	D	81.5%	1.287	-	0.999	-	0.442	0.042
G7+S&P	C - C	0.0%	-	-	-	-	-	-
(61,26)	C - D	10.2%	1.229	0.628	0.982	0.890	0.534	0.040
	D - D	3.4%	1.391	1.154	0.984	0.887	0.785	0.042

Table 3: MS-ARFIMA$(0,d,0)$ breaks in mean and d results

Data (T,N)	Conv.	Perc.	Avg. d_1	Avg. d_2	Avg. P_{11}	Avg. P_{22}	Avg. μ_1	Avg. μ_2	Avg. σ
	C	3.7%	0.892	-	0.999	-	0.781	-	0.056
	D	21.7%	1.218	-	0.999	-	0.728	-	0.054
1930	C - C	3.7%	-0.15	0.923	0.867	0.975	0.338	0.31	0.047
(81,36)	C - D	37.6%	0.014	1.355	0.899	0.972	0.376	0.351	0.049
	D - D	33.3%	1.227	1.613	0.967	0.887	0.183	0.169	0.049
	C	3.0%	0.92	-	0.999	-	0.752	-	0.056
	D	29.5%	1.25	-	0.999	-	0.544	-	0.048
1940	C - C	2.9%	-0.76	0.93	0.845	0.978	0.391	0.459	0.043
(71,36)	C - D	37.1%	-0.484	1.382	0.893	0.978	0.386	0.342	0.042
	D - D	27.5%	1.255	1.734	0.96	0.892	0.118	0.154	0.045
	C	5.1%	0.911	-	0.995	-	0.824	-	0.062
	D	58.6%	1.276	-	0.997	-	0.702	-	0.055
1950	C - C	0.6%	0.337	0.908	0.909	0.958	0.372	0.408	0.058
(61,95)	C - D	20.3%	-0.01	1.336	0.911	0.977	0.48	0.367	0.055
	D - D	15.4%	1.227	1.539	0.953	0.917	0.405	0.146	0.054
	C	3.7%	0.918	-	0.999	-	0.927	-	0.038
	D	61.7%	1.281	-	0.999	-	0.194	-	0.033
Europe + G7	C - C	0.0%	-	-	-	-	-	-	-
(61,25)	C - D	21.3%	-0.356	1.386	0.916	0.98	0.18	0.249	0.028
	D - D	13.3%	1.248	1.614	0.952	0.91	0.18	0.122	0.031
	C	7.4%	0.886	-	0.923	-	0.573	-	0.036
	D	68.0%	1.322	-	0.97	-	0.142	-	0.034
Europe	C - C	0.0%	-	-	-	-	-	-	-
(61,22)	C - D	15.2%	0.038	1.338	0.884	0.989	0.182	0.286	0.028
	D - D	9.5%	1.243	1.557	0.949	0.906	-0.06	-0.129	0.029
	C	2.1%	0.931	-	0.999	-	1.029	-	0.046
	D	65.6%	1.306	-	0.999	-	0.582	-	0.04
Europe + S&P	C - C	0.0%	-	-	-	-	-	-	-
(61,37)	C - D	17.3%	-0.169	1.39	0.906	0.981	0.239	0.285	0.035
	D - D	15.0%	1.242	1.528	0.957	0.932	0.218	0.102	0.035
	C	3.1%	0.942	-	0.999	-	0.873	-	0.045
	D	69.8%	1.321	-	0.999	-	0.511	-	0.041
G7 + S&P	C - C	0.0%	-	-	-	-	-	-	-
(61,26)	C - D	13.5%	-0.228	1.339	0.904	0.983	0.305	0.378	0.036
	D - D	13.5%	1.276	1.53	0.956	0.904	0.379	0.289	0.035

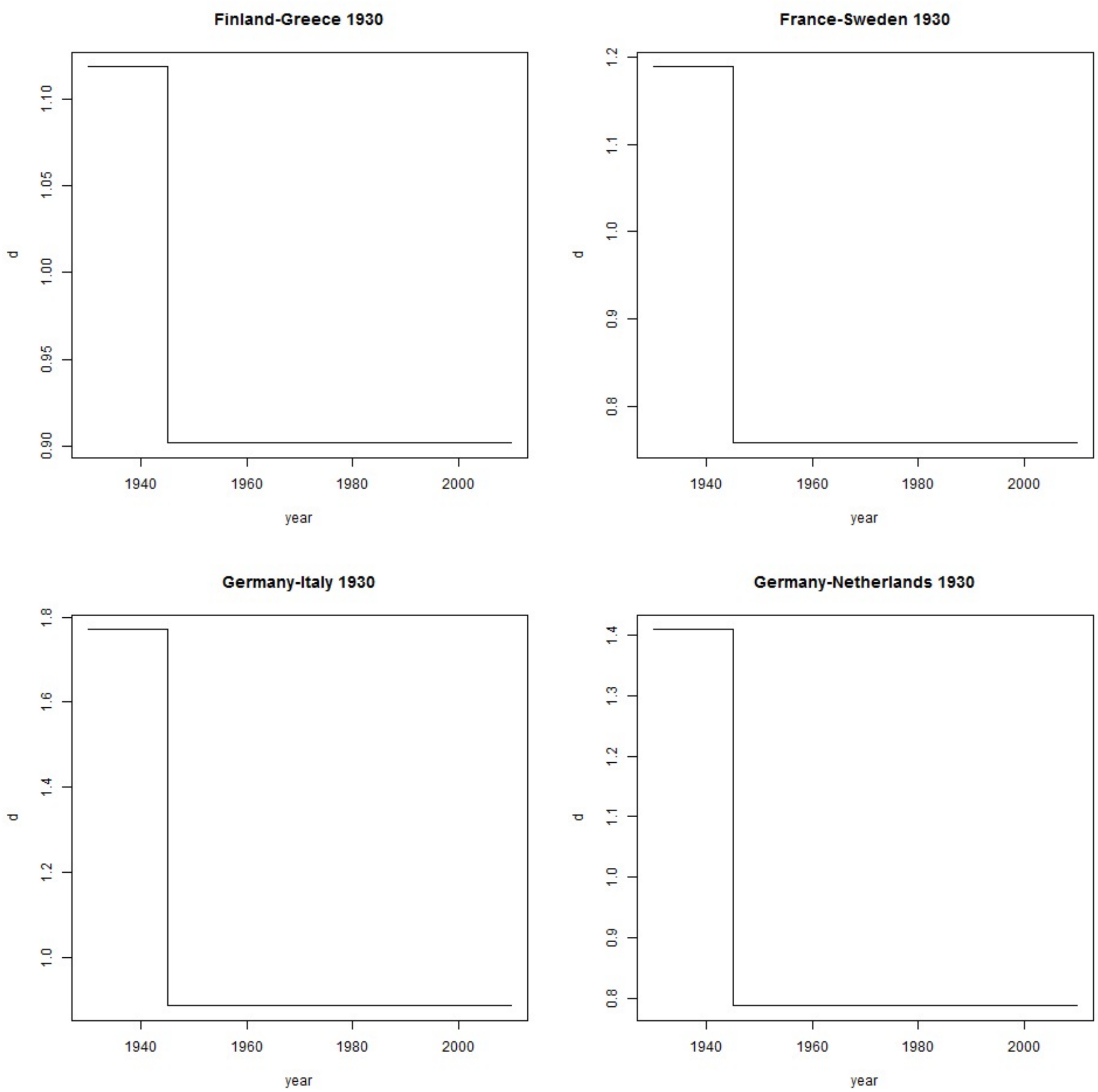

Figure 1: Europe Case, only d Regime Switching

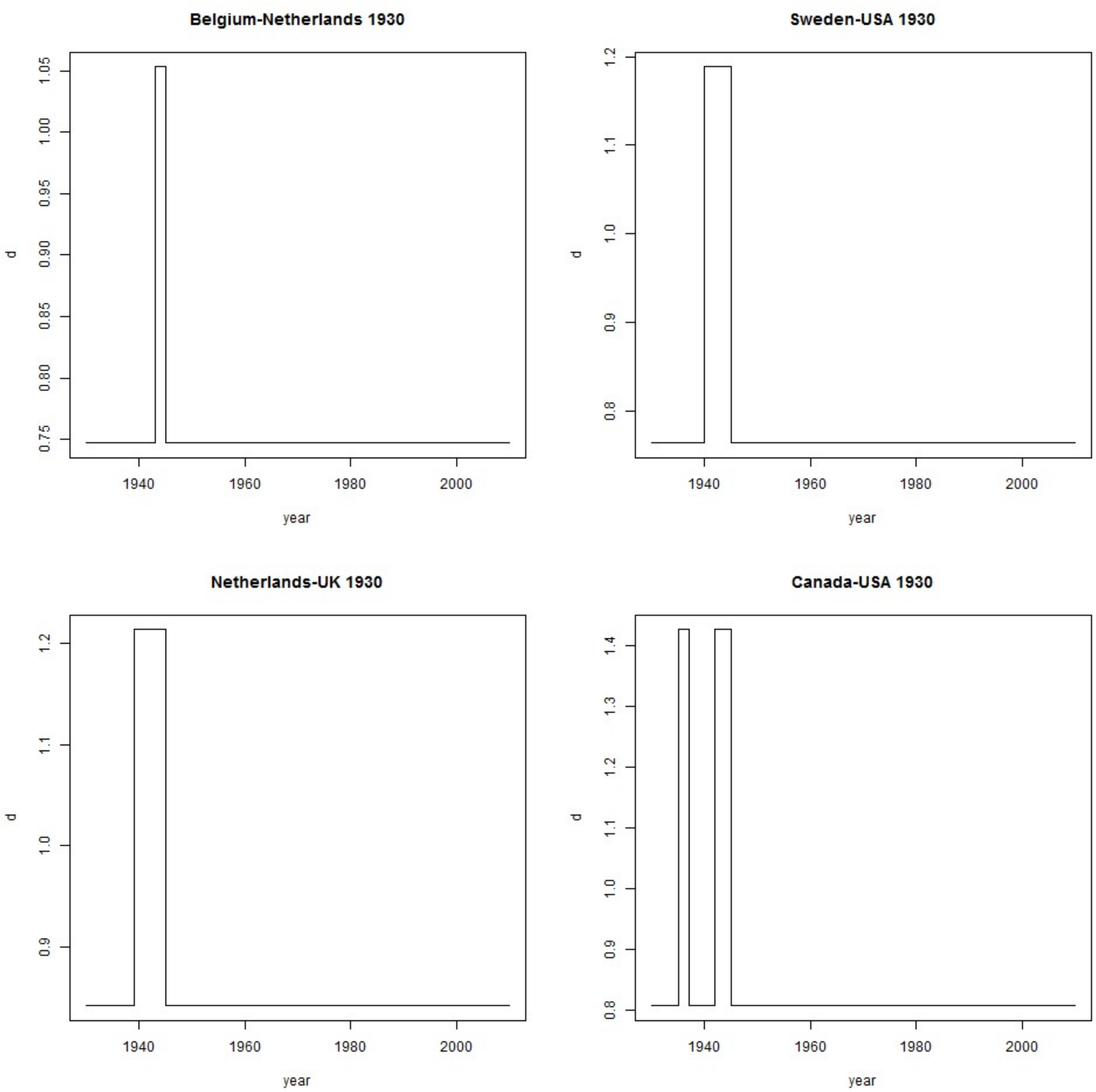

Figure 2: Developed Countries, only *d* Regime Switching

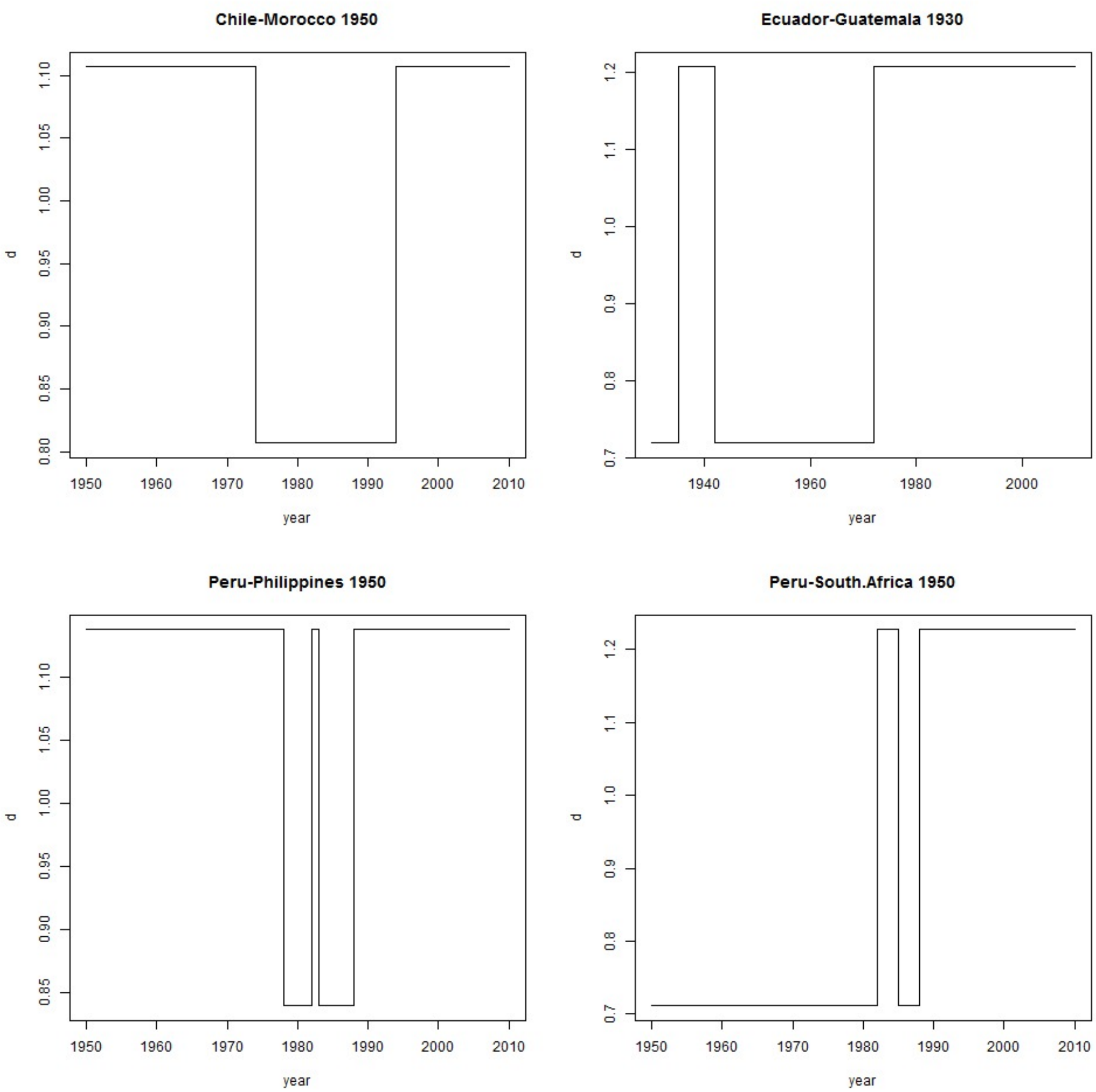

Figure 3: Developing Countries, only d Regime Switching

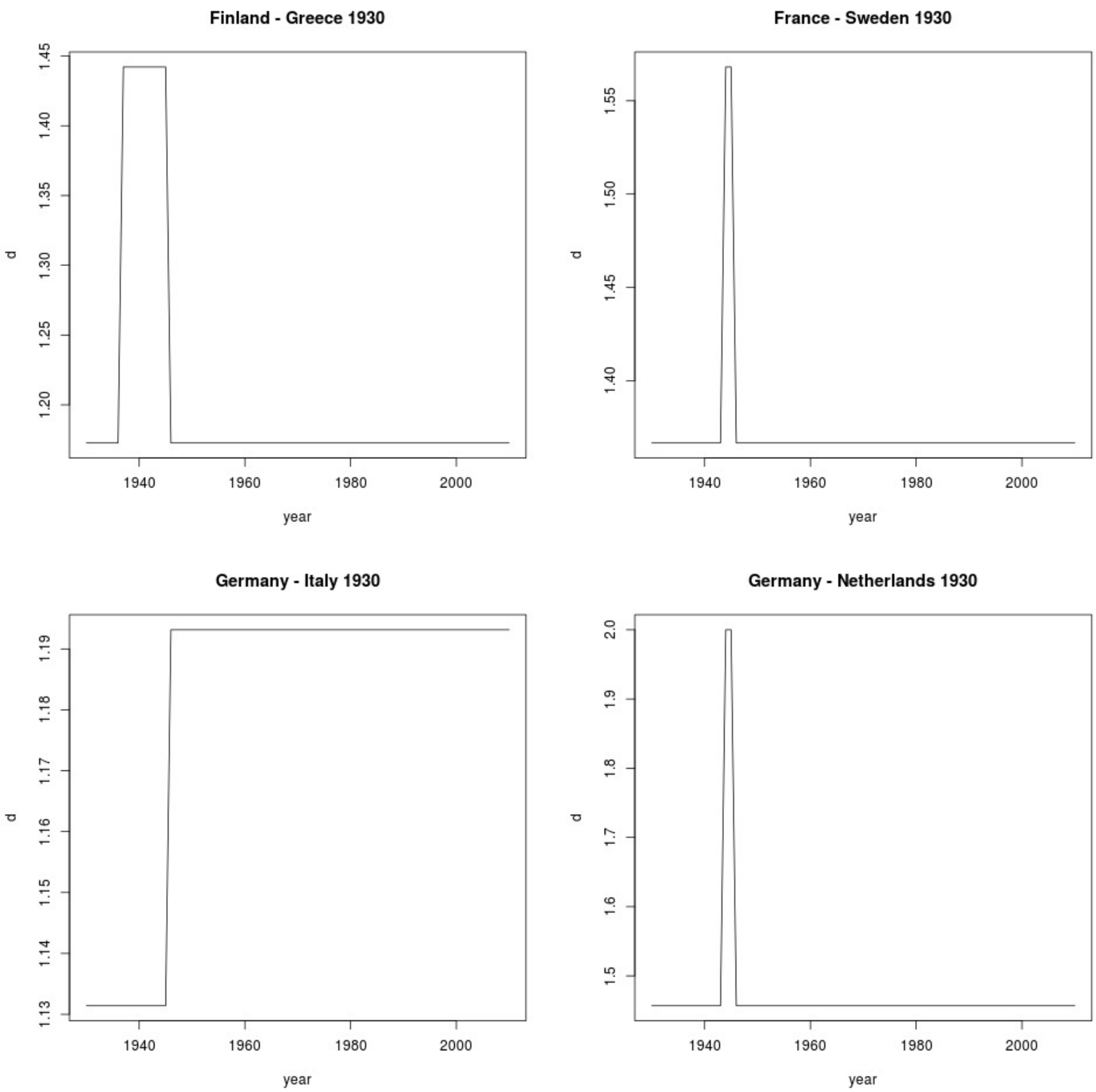

Figure 4: Europe Case, both μ and d Regime Switching

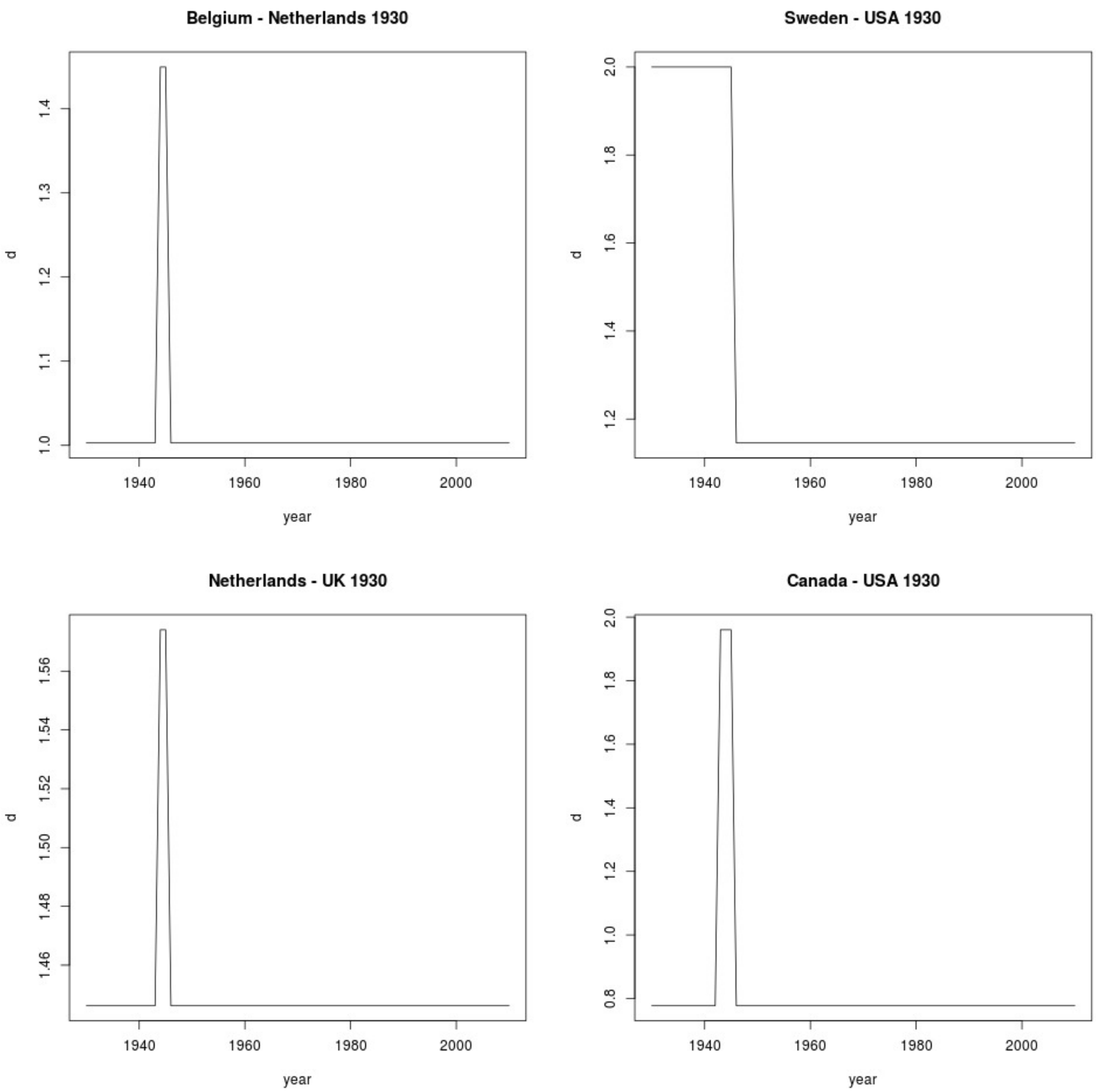

Figure 5: Developed Countries, both μ and d Regime Switching

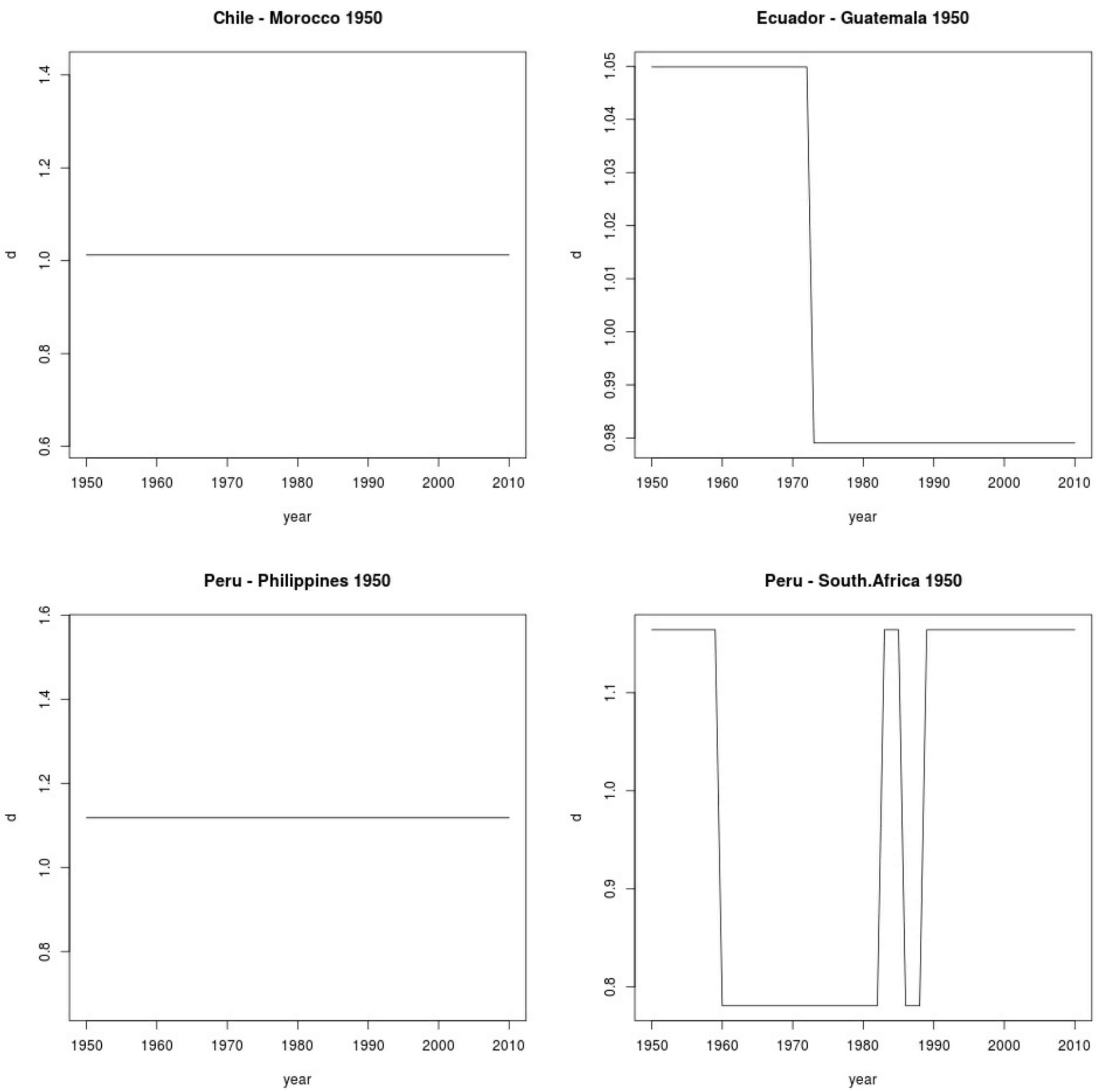

Figure 6: Developing Countries, both μ and d Regime Switching

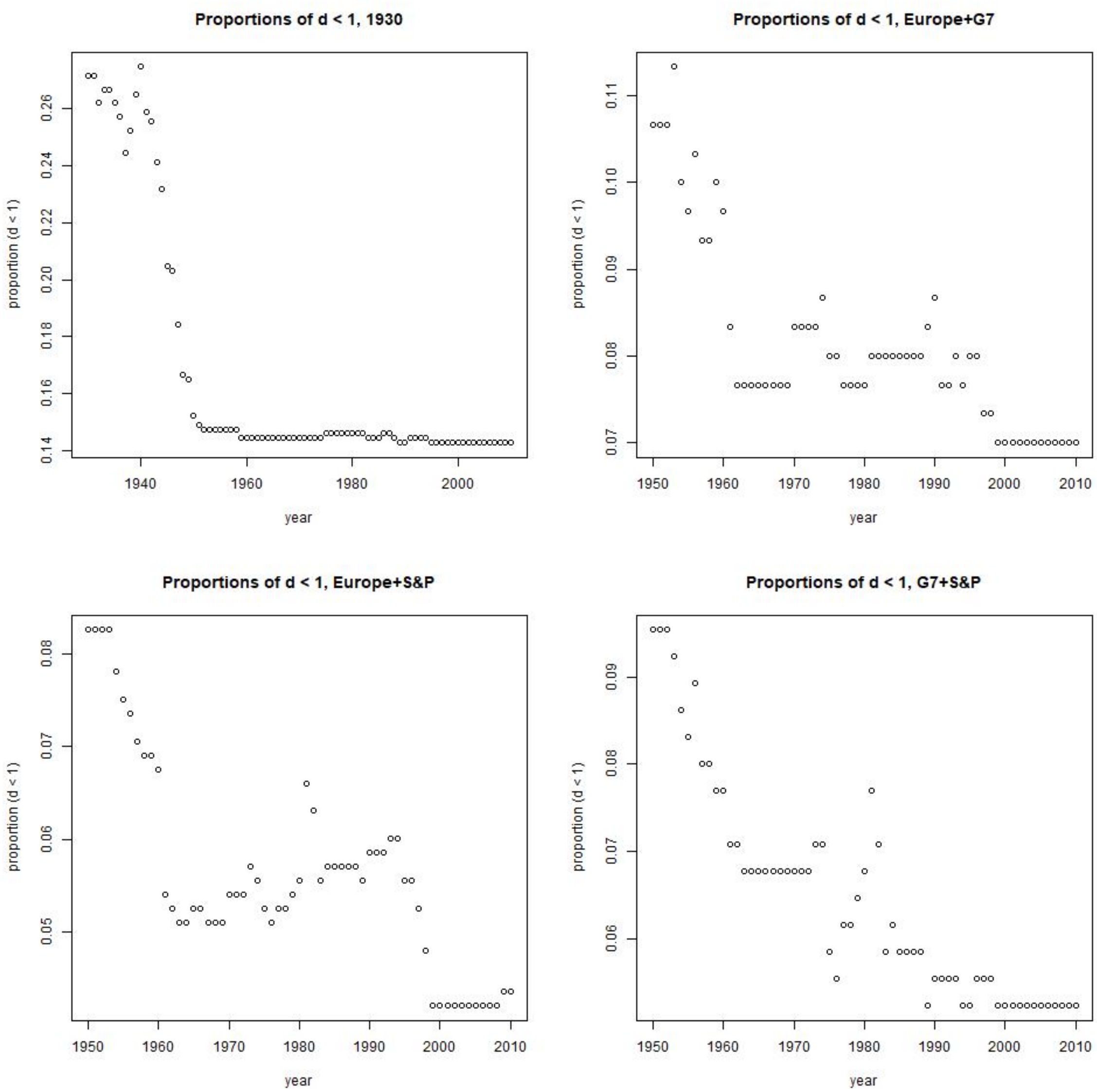

Figure 7: Proportion of convergent country pairs ($d < 1$) over time (only d Regime Switching

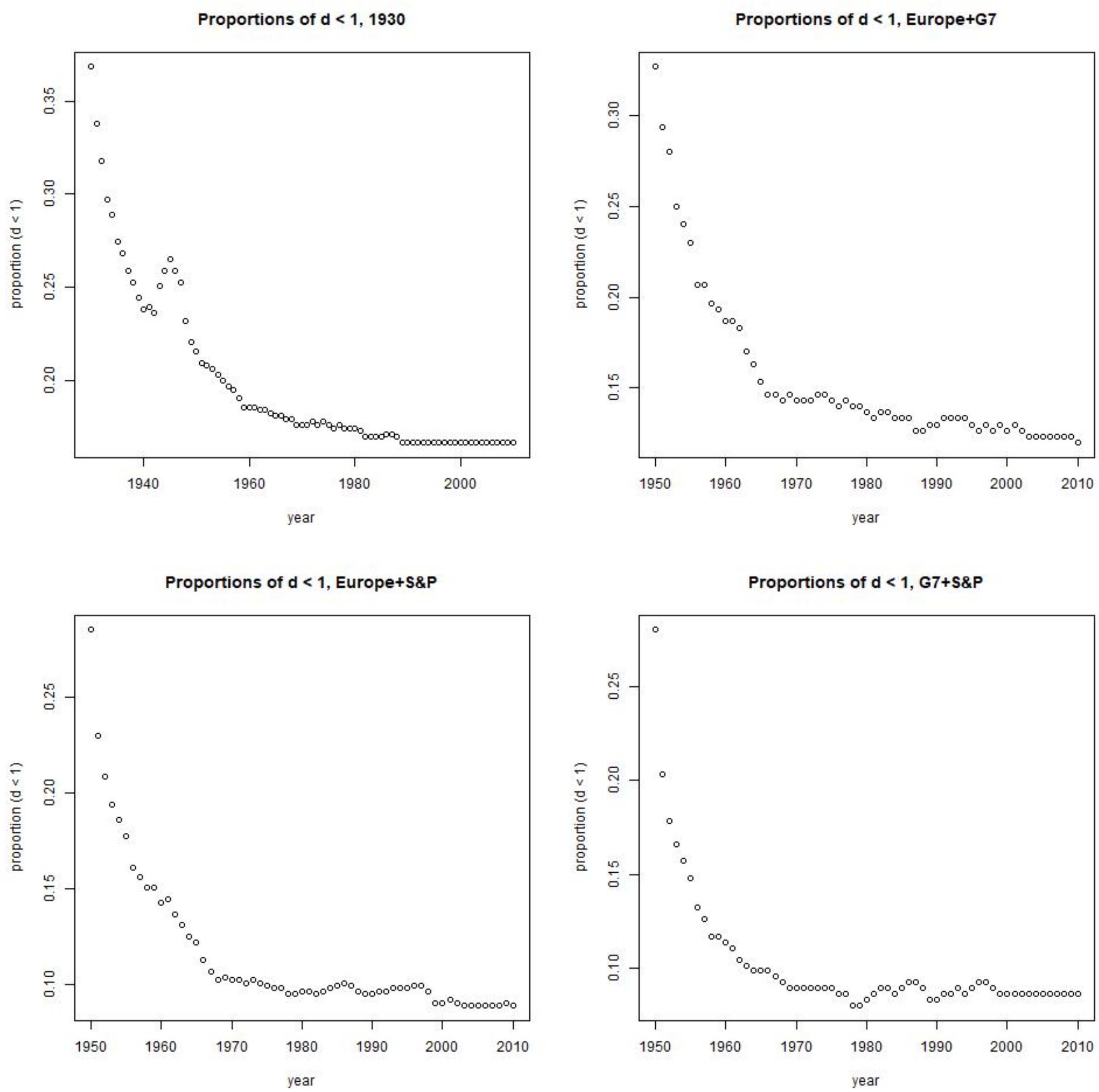

Figure 8: Proportion of convergent country pairs ($d < 1$) over time (both μ and d Regime Switching)

References

Azariadis, C. (1996). The economics of poverty traps part one: complete markets. *Journal of economic growth 1*(4), 449–486.

Azariadis, C. and A. Drazen (1990). Threshold externalities in economic development. *The Quarterly Journal of Economics 105*(2), 501–526.

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

Bernard, A. B. and S. N. Durlauf (1995). Convergence in international output. *Journal of applied econometrics 10*(2), 97–108.

Diebold, F. X. and A. Inoue (2001). Long memory and regime switching. *Journal of econometrics 105*(1), 131–159.

Dufrénot, G., V. Mignon, and T. Naccache (2012). Testing catching-up between the developing countries:growth resistance and sometimes growth tragedy. *Bulletin of Economic Research 64*(4), 470–508.

Durlauf, S. N., P. A. Johnson, and J. R. Temple (2005). Growth econometrics. *Handbook of economic growth 1*, 555–677.

Haldrup, N. and M. Nielsen (2005). A regime switching long memory model for electricity prices, forthcoming in journal of econometrics.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, 357–384.

Johnson, P. and C. Papageorgiou (2017). What remains of cross-country convergence? *Journal of Economic Literature* (forthcoming).

Kerekes, M. (2012). Growth miracles and failures in a markov switching classification model of growth. *Journal of Development Economics 98*(2), 167–177.

Li, Q. and D. Papell (1999). Convergence of international output time series evidence for 16 oecd countries. *International review of economics & finance 8*(3), 267–280.

Michelacci, C. and P. Zaffaroni (2000). (fractional) beta convergence. *Journal of Monetary Economics 45*(1), 129–153.

Özkan, H., T. Stengos, and E. Yazgan (2016). Markov regime switching in mean and in fractional integration parameter. *Communications in Statistics-Simulation and Computation* (forthcoming).

Perron, P. et al. (2006). Dealing with structural breaks. *Palgrave handbook of econometrics 1*, 278–352.

Pesaran, H. M. (2007). A pair-wise approach to testing for output and growth convergence. *Journal of Econometrics 138*(1), 312–355.

Phillips, P. C. and D. Sul (2007a). Some empirics on economic growth under heterogeneous technology. *Journal of Macroeconomics 29*(3), 455–469.

Phillips, P. C. and D. Sul (2007b). Transition modeling and econometric convergence tests. *Econometrica 75*(6), 1771–1855.

Romer, P. M. (1994). The origins of endogenous growth. *The journal of economic perspectives 8*(1), 3–22.

Stengos, T., B. S. Thompson, and X. Wu (2009). The evolution of the conditional joint distribution of life expectancy and per capita income growth. In *Nonparametric Econometric Methods*, pp. 171–191. Emerald Group Publishing Limited.

Stengos, T. and M. Yazgan (2014a). Persistence in real exchange rate convergence. *Studies in Nonlinear Dynamics and Econometrics 18*(1), 73–88.

Stengos, T. and M. E. Yazgan (2014b). Persistence in convergence. *Macroeconomic Dynamics 18*(04), 753–782.

Tsay, W.-J. and W. K. Härdle (2009). A generalized arfima process with markov-switching fractional differencing parameter. *Journal of Statistical Computation and Simulation 79*(5), 731–745.

Viterbi, A. (1967). Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE transactions on Information Theory 13*(2), 260–269.

Wu, X., A. Savvides, and T. Stengos (2014). The global joint distribution of income and health. In *Recent Advances in Estimating Nonlinear Models*, pp. 249–279. Springer.