Regime Switching with Structural Breaks in Output Convergence: Further Results

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

June, 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 used by Hamilton (1989) 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. Our approach is based on estimation of d alone and it does not rest on testing in order to avoid problems of size distortions that may plague testing based approaches pursued in the literature, see Pesaran (2007) and Stengos and Yazgan (2014a).

JEL Classification: C12, C22

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 between different economies has been an area of intensive research in the late thirty years. Durlauf et al. (2005) provide a comprehensive survey of the recent developments and evidence regarding the convergence hypothesis. 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. 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. Dufrénot et al. (2012) 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) and Datta (2003) among others. 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.

Ideally, one would like to examine output gaps that may transition between different regimes (convergence and divergence) by directly observing these transitions and unravelling the movements between convergent and divergent regimes. 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. Tsay and Härdle (2009) introduced 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 disentangle the impact of long memory dependence on the estimates of the latent regime parameters in the MS–ARFIMA framework. However, their analysis although allows for regime switching in the long memory parameter, only concentrated 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 et al. (2016) within the Tsay and Härdle (2009) set up allowed for breaks both in the mean and long memory parameter. Using Monte Carlo simulations 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 a recent paper, Beylunioglu, Stengos and Yazgan (2018) 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. As existing methods used in the estimation of standard MS models, such as the EM algorithm used by Hamilton (1989), are no longer appropriate, they make use of the Viterbi algorithm to estimate the long memory MS model used by Tsay and Härdle (2009). However, they only 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. The convergence (or lack of it) for the group of countries under consideration is determined by the overall proportion of pairs showing stationary behavior. They also examine the path of the d parameter over time to observe directly the movement between the two regimes (the transition from convergence to divergence) for given pairs. However, since they only examine two regimes they are in effect considering a simple dichotomous $I(1)$ versus $I(0)$ set up.

In the present paper we extend the Beylunioğlu et al. (2018) approach to allow for three regimes, something that would better match a long memory framework with structural breaks, than Beylunioğlu et al. (2018) did in their work. By examining the path of d parameter over time, enables us to observe the transitions between convergent, mean reverting and non-convergent behavior directly, we find that converging behavior is diminishing over time and lack of convergence is the dominant force.

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

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

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 (2014a)). 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.

The pairwise method developed by Pesaran (2007) can offer a possible remedy to both of the above difficulties. 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 ($H_0 : Z_t \sim I(1)$) 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. 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. To summarize, rejection rates higher than a given significance level in a given application would imply evidence against the non-convergence hypothesis. On the other hand, rejection rates lower or close to the employed significance level will provide evidence for the non-rejection (validity) of the non-convergence hypothesis. However, this approach is also subject to serious 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 is a fractional integration parameter $-0.5 < d \leq 1$ there is a much richer classification. One can distinguish between the different convergence cases that are implied by the processes above. Different values of d will define different types of convergence and we enumerate these different convergence cases below. For different values of d, we have the following three states that we will also use in our classification approach.

State 1 (stationary) : This combines the cases where $0 \leq d < 0.5$. This is the case of a short memory process, where there is "fast catching-up" or "short memory catching-up" when $-0.5 < d \leq 0$ and 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, when $0 < d \leq 0.5$.

State 2 (mean-reverting) : $0.5 \leq d < 1$, that is 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.

State 3: (divergent) $d \geq 1$, where we have 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.

In the previous literature with a long memory approach, 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. 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). Dufrénot et al. (2012) and Stengos and Yazgan (2014a) 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 belongs to. However, there are serious issues of low power and size distortions that may affect the above sequential testing methodology, something that is inherent in any testing-based approach.

To avoid the above potentially serious problems that may affect any sequential testing approach 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 converging (both rapid and mean reverting) and non-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 within the three states outlined above 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 and those that reflect rapid convergent and mean reverting processes. 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 state position $0 \leq d < 0.5$, $0.5 \leq d < 1$ and $d \geq 1$ and the presence of convergence is determined from the overall results of these groups.

More precisely, 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 in each of the three states. 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. The approach that we follow simply centers on the value of d and we ignore the fact that d may bifurcate between regimes as we treat the two values as separate estimates. 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 that we use.

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

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

where $\varepsilon_t \sim N(0, \sigma^2)$ $I.I.D$, $L(\cdot)$ is the lagging operator and d_{st} is long memory parameter. Notice that we omit i and j in order to simplify the notation σ is constant for each pairs, allowing d_{s_t} and μ_{s_t} to vary over time and among states.

For T observations, we assume a hidden markov process $\{s_t\}_{t=1}^{T}$ driving gap series where $s_t \in \{D, M, S\}$ implies the regime at time t, is either divergent, mean reverting or stationary state respectively. We also assume that the states are satisfying Markov property according to the following transition matrix,

$$\mathcal{P} = \begin{bmatrix} P_{DD} & P_{DM} & P_{DS} \\ P_{MD} & P_{MM} & P_{MS} \\ P_{SD} & P_{SM} & P_{SS} \end{bmatrix}$$

where $P_{ij} = \mathbb{P}(s_t = j \mid s_{t-1} = i)$ and $P_{iD} + P_{iM} + P_{iS} = 1$ for $i = D, M, S$.

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

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 $\zeta = \{\mu_D, \mu_M, \mu_S, \sigma, d_D, d_M, d_S, \mathbb{P}_{DD}, \mathbb{P}_{MM}, \mathbb{P}_{SS}, \mathbb{P}_{DM}, \mathbb{P}_{MD}, \mathbb{P}_{SD}\}$. However in real case we do not observe the state series $s_{1:T} = \{s_1, s_2, \ldots, s_T\}$ but $Z_{1:T}$ and need to estimate the most likely path of states driving the outcomes. The estimation is computationally complex since it searches the only most likely path out of 3^T possibilities; more clearly we need to evaluate all paths to determine the most likely one. To reduce this complexity, we will adopt MS-ARFIMA(p, d, q) model proposed by Tsay and Härdle (2009) which operates 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 chooses the most likely path passing through each node. Since we have three regimes at each step, we have three best paths (survivors) that ends in D, M and S respectively. To exemplify, for $t = 2$, the algorithm evaluates all 9 paths conditioned to results of previous step and keeps three survivors; thus for the next step, $t = 3$, we end in 9 possible paths with three prospective survivors. This evaluation continues to the last step where one best route having the largest likelihood 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 limiting probabilities. Thus, at each step we evaluate the paths passing through nodes and one path for each $s_t \in \{D, M, S\}$ by $s^*_{1:t-1} = \arg M(s_t)$. Once the evaluation is completed, the algorithm returns three processes with their likelihood estimations among which the most likely is chosen.

In (5), only term that links observations and state processes is $\mathbb{P}(Z_t|s_t)$ over which the likelihood is calculated for given $s_{1:T}$ and $Z_{1:T}$. To fit the likelihood into recursive form, Tsay and Härdle (2009) join Durbin-Levinson recursion and rearranges likelihood function for a given path as,

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

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, estimation of ζ amongst infinite set of parameters requires constrained optimization procedure to satisfy the constraint $P_{iD} + P_{iM} + P_{iS} = 1$ which can only be restricted externally. To perform our analysis, we used R platform and operated constrOptim function in base library for numerical optimization*.

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.

Data and Empirical Findings

We update the Maddison data set that was *used in Stengos and Yazgan (2014)*. The Maddison data consist of annual GDP per capita data for different time periods and for different country groups depending on the

*The computations were made with a 4 cores/8 threads i7 Linux computer. For the 1930 data with 630 pairs and 81 time-points each, it took 4520 seconds (~72 minutes).

period. The period covering 1950 to 2010 for example consists of 141 countries[†]. The country coverage is given in Table 2 for all groups considered. We considered four different data sets, one starting from 1930 and three being 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[‡].

Table 1: MS-ARFIMA$(0, d, 0)$ breaks in both μ and d results

Data	Type	Percentage	Avg. d_1	Avg. d_2	Avg. d_3	Avg. P_{11}	Avg. P_{22}	Avg. P_{33}	Avg. μ_1	Avg. μ_2	Avg. μ_3	Avg. σ
1930	SSS	0.00%										
	MMM	5.08%	0.884	0.894	0.905	0.967	0.977	0.993	0.918	0.916	0.903	0.058
	DDD	48.89%	1.203	1.304	1.405	0.953	0.930	0.915	0.545	0.557	0.541	0.049
	SM	0.00%										
	MD	11.59%	0.788	1.422		0.868	0.961		0.276	0.169		0.050
	SD	6.35%	-0.232	1.322		0.860	0.989		0.504	0.582		0.049
	SSM	0.00%										
	SMM	0.16%	-1.459	0.749	0.927	0.600	0.600	0.987	0.685	0.132	-0.039	0.030
	MMD	3.65%	0.746	0.877	1.377	0.713	0.796	0.941	0.122	0.183	0.152	0.041
	MDD	12.22%	0.747	1.281	1.565	0.827	0.851	0.824	0.036	0.052	0.187	0.043
	SSD	0.32%	-0.880	-0.004	1.189	0.831	0.737	0.987	0.133	0.138	0.225	0.058
	SDD	7.78%	0.003	1.244	1.586	0.785	0.886	0.865	0.299	0.314	0.251	0.046
	SMD	3.97%	-0.266	0.811	1.421	0.761	0.786	0.943	0.389	0.247	0.267	0.045
Europe+S&P	SSS	0.00%										
	MMM	1.95%	0.931	0.931	0.931	0.999	0.999	0.999	0.939	0.939	0.939	0.046
	DDD	78.08%	1.298	1.329	1.359	0.981	0.979	0.976	0.530	0.519	0.511	0.039
	SM	0.00%										
	MD	9.91%	0.791	1.468		0.886	0.973		0.287	0.136		0.037
	SD	5.26%	-0.093	1.399		0.859	0.987		0.510	0.523		0.038
	SSM	0.00%										
	SMM	0.00%										
	MMD	0.30%	0.576	0.773	1.531	0.887	0.759	0.965	0.184	0.176	-0.021	0.018
	MDD	2.10%	0.806	1.333	1.513	0.845	0.857	0.865	0.104	0.118	0.031	0.034
	SSD	0.15%	-0.534	0.374	1.437	0.802	0.895	0.999	0.075	0.152	0.428	0.022
	SDD	2.10%	0.279	1.242	1.565	0.837	0.809	0.888	-0.020	-0.016	0.066	0.033
	SMD	0.15%	-0.554	0.861	1.413	0.741	0.808	0.985	0.014	0.017	0.666	0.034
G7+Europe	SSS	0.00%										
	MMM	3.00%	0.914	0.914	0.914	0.999	0.999	0.999	0.858	0.858	0.858	0.042
	DDD	76.67%	1.276	1.312	1.346	0.976	0.974	0.969	0.207	0.208	0.211	0.033
	SM	0.00%										
	MD	8.67%	0.771	1.520		0.858	0.956		0.031	0.095		0.029
	SD	6.67%	-0.210	1.456		0.846	0.988		0.276	0.393		0.030
	SSM	0.00%										
	SMM	0.00%										
	MMD	0.67%	0.658	0.723	1.515	0.767	0.726	0.961	0.095	0.145	0.149	0.016
	MDD	1.67%	0.836	1.369	1.558	0.805	0.911	0.818	0.173	0.128	0.136	0.024
	SSD	0.67%	-0.240	0.401	1.445	0.751	0.902	0.978	0.187	0.146	0.340	0.019
	SDD	1.67%	-0.210	1.332	1.636	0.792	0.858	0.854	-0.132	-0.053	0.007	0.022
	SMD	0.33%	0.144	0.689	1.277	0.701	0.600	0.980	0.399	0.350	0.502	0.024
G7+S&P	SSS	0.62%	0.082	0.144	0.206	0.847	0.777	0.708	1.019	1.014	1.010	0.042
	MMM	8.00%	0.850	0.854	0.858	0.831	0.832	0.834	0.305	0.298	0.291	0.045
	DDD	68.92%	1.338	1.363	1.388	0.936	0.936	0.933	0.494	0.493	0.482	0.039
	SM	0.92%	0.181	0.778		0.690	0.754		0.235	-0.524		0.037
	MD	13.85%	0.754	1.364		0.810	0.957		0.186	0.404		0.042
	SD	2.46%	0.317	1.470		0.736	0.920		0.105	0.170		0.036
	SSM	0.00%										
	SMM	0.00%										
	MMD	1.23%	0.685	0.953	1.344	0.845	0.784	0.996	0.004	0.001	0.204	0.041
	MDD	2.77%	0.763	1.323	1.598	0.790	0.927	0.844	0.113	-0.007	-0.026	0.040
	SSD	0.00%										
	SDD	1.23%	-0.163	1.178	1.454	0.787	0.843	0.901	0.454	-0.211	0.508	0.034
	SMD	0.00%										

[†]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.

[‡]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.

Table 1 displays the summary of results for the different datasets. The outcomes are displayed in nine groups indicating the different states and the transition between states that are considered in our analysis. We examine three states, stationary convergent (S), mean-reverting (M) and divergent (D) according to the value of d that falls in the corresponding state as defined above. The other cases denote the transition states between S, M and D. In that case, $D - S$ stands for the cases of transitions from divergence to stationary convergence, moving from d estimates such that $d_D \geq 1.$ to $d_S \leq 0.5$. The case $S - S$ denotes the transition from converging to converging behavior, while $D - D$ the case of for diverging to diverging behavior, while $M - S$ the transition from mean reverting to stationary catching up and $D - M$ from divergence to mean-reverting behavior.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 (2014), where no significant convergence is detected. However, one main difference is that we can now identify the type of convergence that may exist to be of the mean reverting variety. In Table 1, the proportion of pairs in the D state are for all the 1950 date sets close to 60% and if one were to add the $D - D$ state the total would be close to 75%. It is only for the 1930 data set that there some more evidence of convergence as the transitions between the different states are more equal. The proportion in the pure S state is almost in all cases zero or close to zero. 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. Clearly, as the convergent single state as well as the transition of a convergent state to another convergent state is so low, the evidence of rapid convergence is nearly absent from our results. However, there is movement between D and M , something that suggests that the type of converging behavior if present is of the mean-reverting variety. This is a result that was not evident from the results of Beylunioğlu et al. (2018) who did not distinguish between rapid catching up and nmean reversion as the type of convergence that they allowed in their analysis. 1940 data.

In order to illustrate the usefuness of our approach in following the movements between the different states over time we also report the plots of d for certain pairs of countries, both developed and developing 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. The figures for developed countries suggest how such pairs as Finland and Norway or Argentina and Guatemala have had estimates of d of their output gap that denote fluctuate between convergent and divergent behavior with d typically becoming large during WWII and then dropping to convergent behavior values for the after the WWII period. For other pairs, such as N. Zealand and Switzerland there have been oscillations that allow for transitions between convergence and divergence.

Finally, since 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 detailed. For Figure-**??** we first obtained d series for each country and calculated the cross-sectional proportion of $d < 1$ for each time point. 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.

3. 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. We extend Beylunioğlu et al. (2018) to allow for three states, stationary rapid convergence, mean reversion and divergence. Even though our results are similar to the findings of Pesaran (2007) and Stengos and Yazgan (2014), in that no sizable convergence is detected, we do find that any convergent behavior is of the mean reverting variety, something that was not seen in the previous literature.

Table 2: Country Groups based on Economic Characteristics and Data Availability for Growth Application

4. Tables and Figures

1930	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 + 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

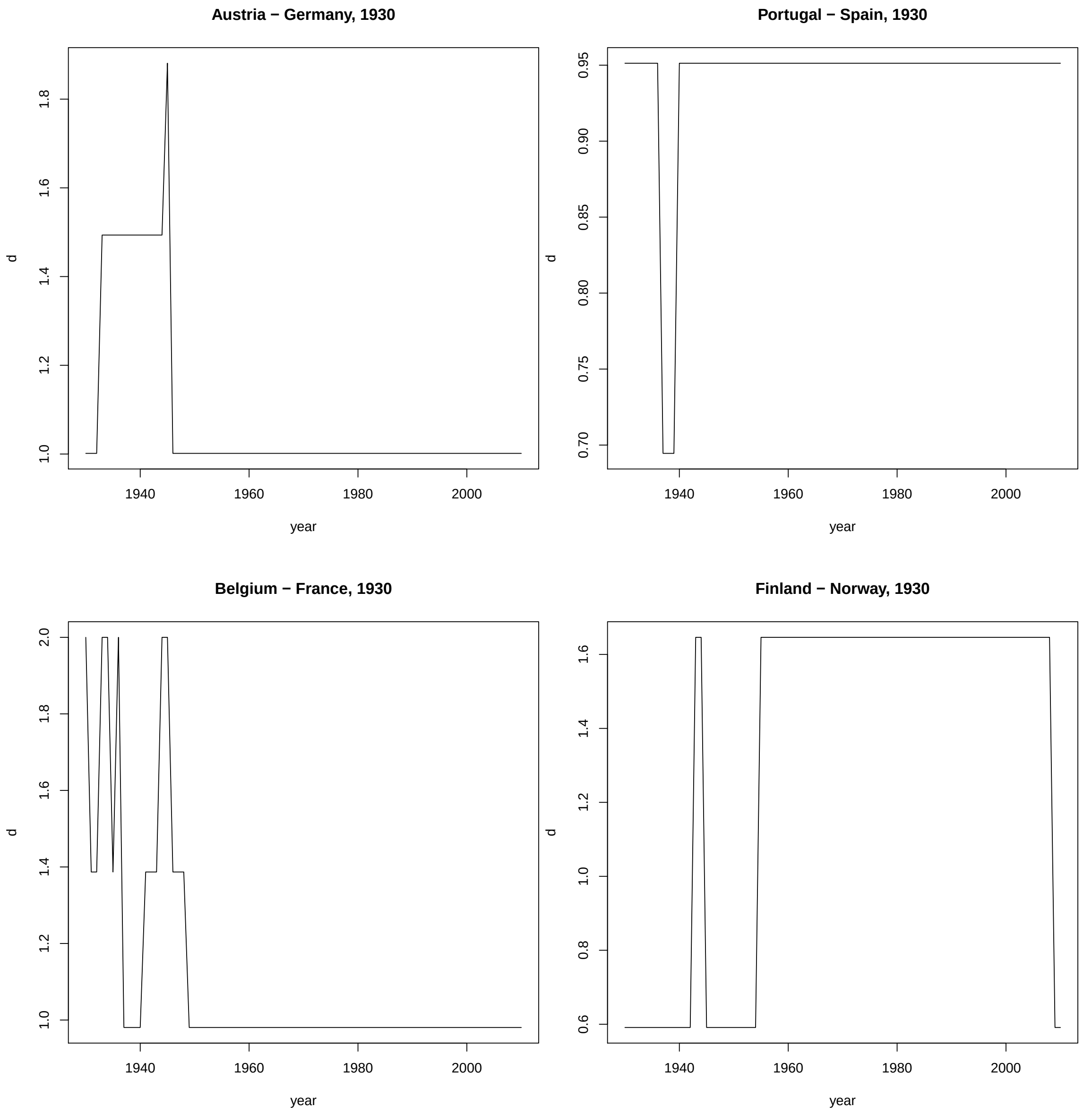

Figure 1: Europe Case

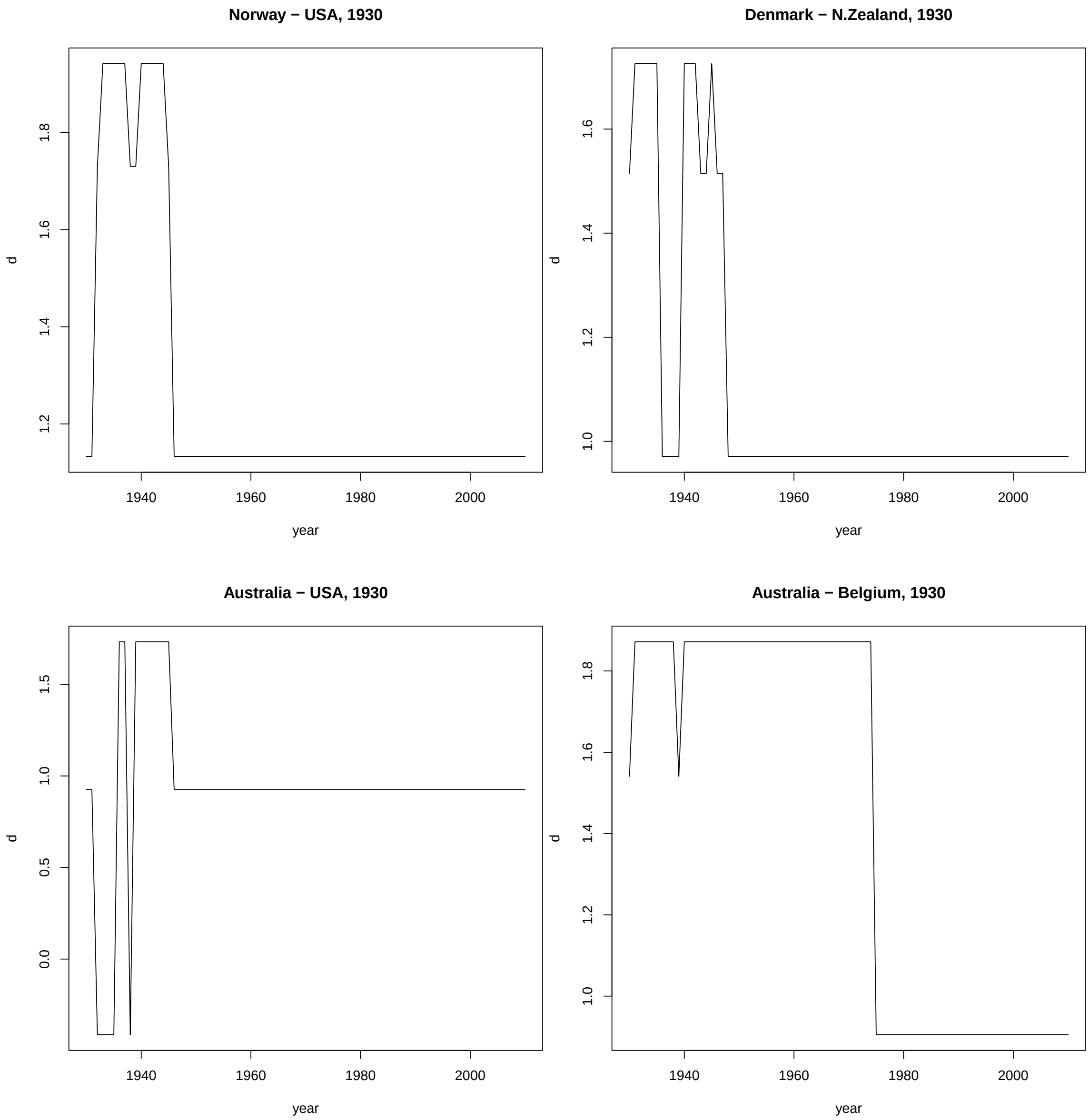

Figure 2: Developed Countries

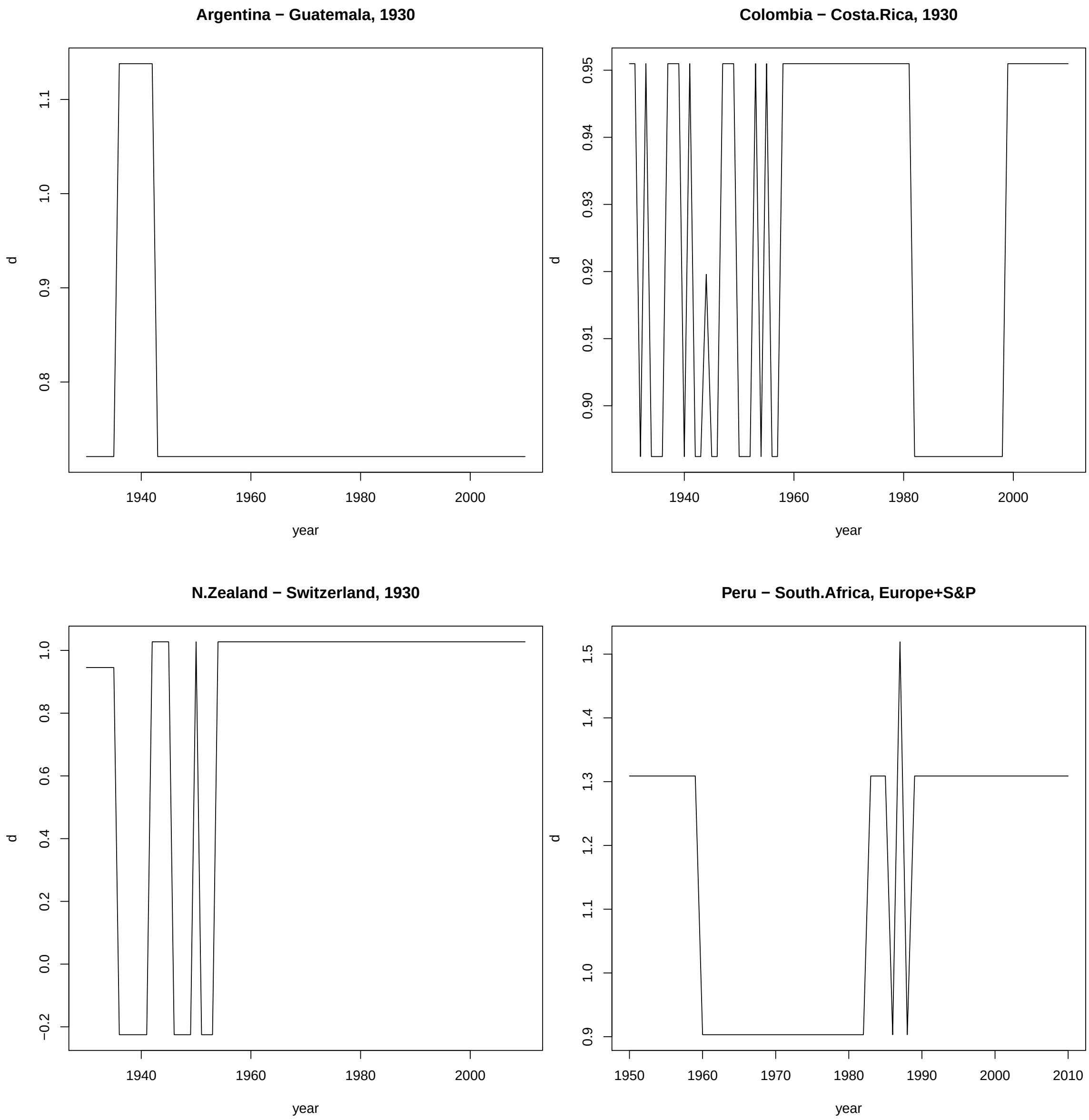

Figure 3: Developing Countries

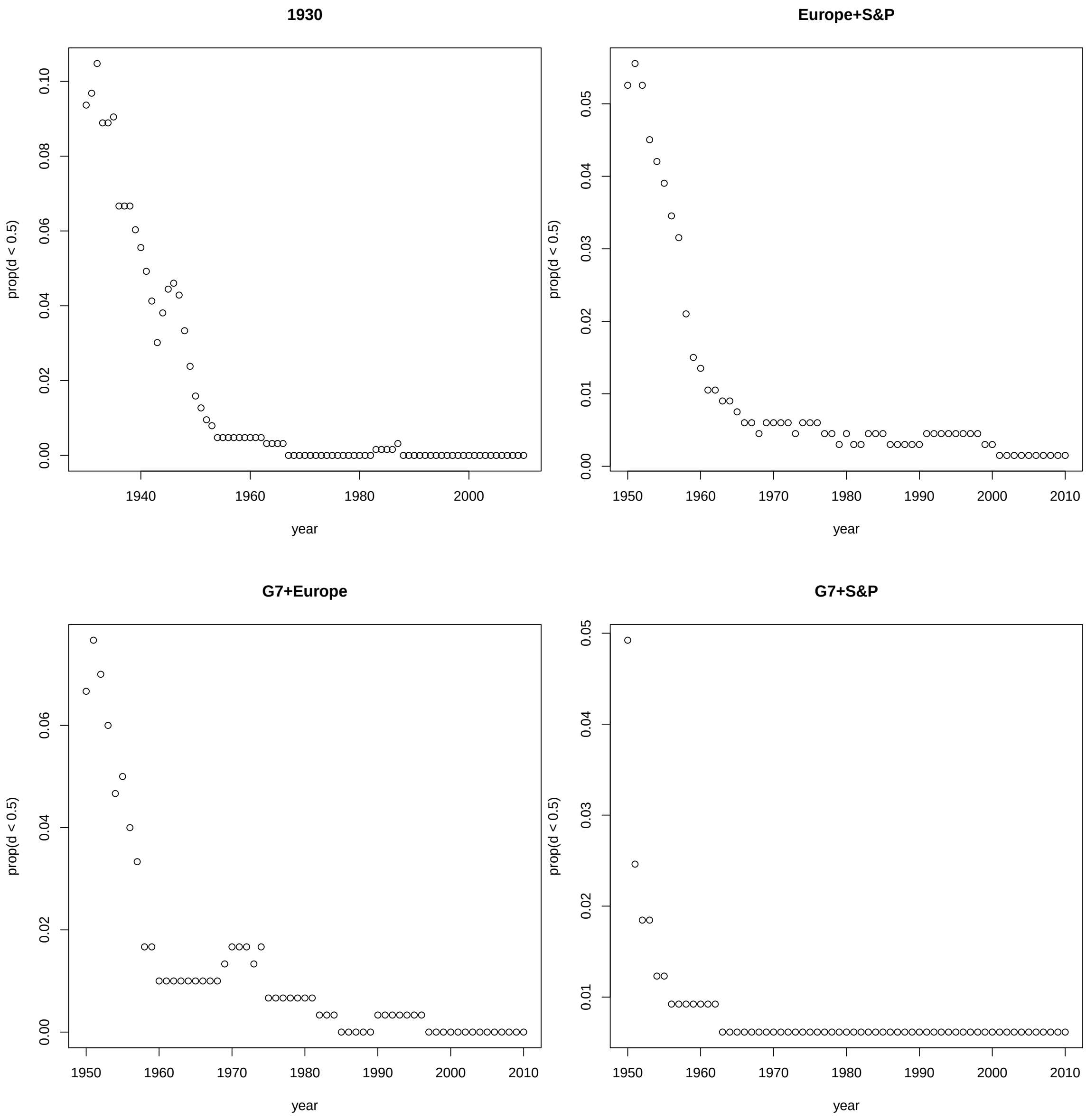

Figure 4: Proportion of stationary over time

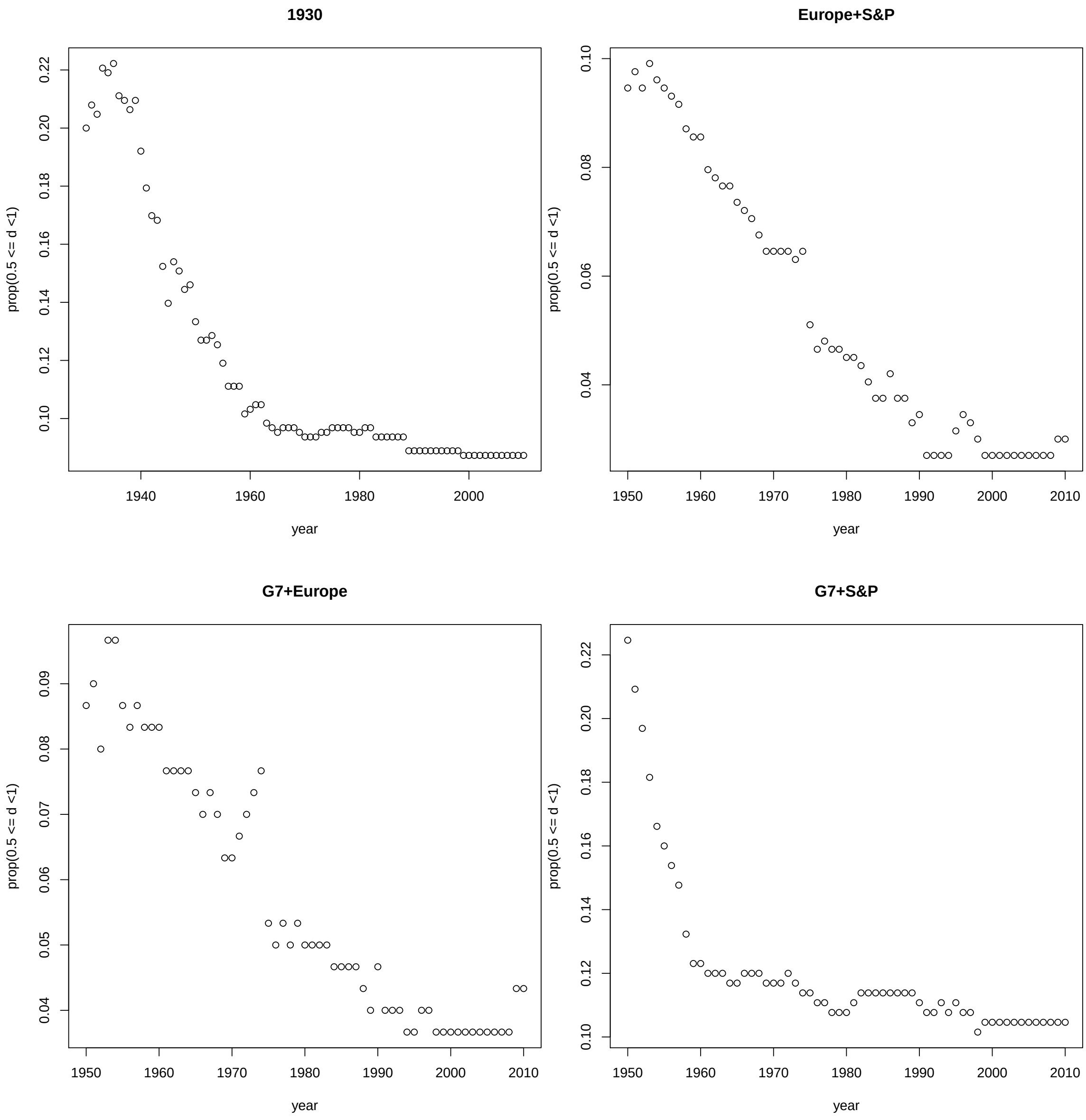

Figure 5: Proportion of mean reverting over time

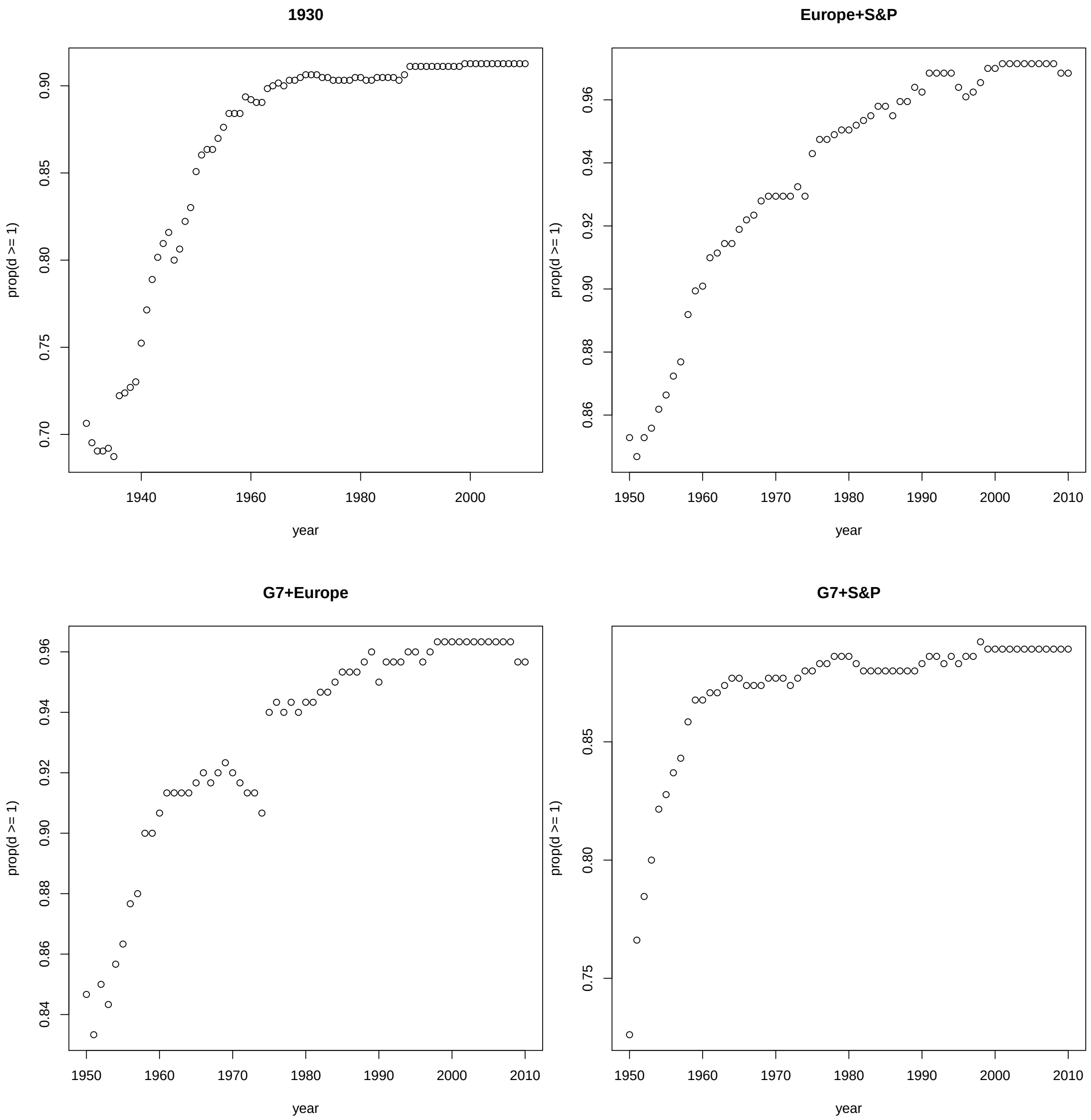

Figure 6: Proportion of non stationary over time

References

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.

Beylunioğlu, F. C., T. Stengos, and M. Yazgan (2018). Regime switching with structural breaks in output convergence. *Studies in Nonlinear Dynamics & Econometrics* (forthcoming).

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.

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.

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.

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

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

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.