Growth Miracles and Failures in a Markov Switching Classification Model of Growth

Monika Kerekes

Growth Miracles and Failures in a Markov Switching Classification Model of Growth

Monika Kerekes[*]

[†]

FU Berlin

June 2009

Abstract

Acknowledging the fact that the growth experience of countries is seldom well described by the average growth rate, this paper aims at identifying countries that are similar in terms of their growth process, thus emphasizing the dynamics of growth rates. To that end, the growth experience of countries is interpreted as a Markov switching process with countries switching between four distinct growth regimes: crisis, stagnation, stable growth, and miracle growth. In the model, different growth patterns arise because countries switch between the growth regimes with different frequencies. In order to account for the distinct dynamics, the traditional Markov switching model is extended by a classification mechanism that endogenously assigns countries exhibiting similar dynamics into the same, and countries exhibiting distinct dynamics into different clusters. Three distinct growth clusters are obtained: the first cluster consists of countries that have achieved relatively fast and steady growth mainly by spending time in the stable and the miracle growth regime. Countries in the second cluster have achieved only moderate growth and often found themselves in stagnation for longer periods. The third cluster might be referred as a growth failure cluster because the countries associated with this cluster have suffered from small growth rates and frequent crises. It appears that developing countries can avoid falling into the growth failure cluster by securing a minimum amount of human capital. In contrast to that, the most distinguishing feature of the countries in the successful growth cluster is their reasonable quality of institutions.

Keywords: Economic Growth, Regime Switching, Latent Class Models
JEL Classification: O11, O57, C23

---

1 Introduction

In recent years the shortcomings of the traditional empirical growth literature, i. e. the cross-country and panel growth regressions, have gained attention. A major problem of growth theories is that they are open-ended, which means that the validity of one causal theory of growth does not invalidate another. Therefore, the specification of growth models and in particular the question which variables to include in growth regressions is difficult to solve. A second problem arises from this openness. Because any growth regression will exclude relevant variables, it is almost impossible to find suitable instruments for endogenous regressors that are both correlated with the regressor and uncorrelated with the omitted growth factors. Moreover, most growth regressions implicitly assume a homogeneous effect of growth regressors across countries, which is quite unlikely given the markedly different country environments.[1] Finally, it has been questioned whether the explanation of average growth rates is appropriate given that the growth rates of developing countries are highly instable and that therefore the results of any growth regression depend to a large extent on the exact time period(s) chosen for the analysis (Pritchett, 2000). The idea of the present study is related to these shortcomings: Instead of focusing on the average growth rate it focuses on the growth process itself. It tries to identify groups of countries that follow similar growth processes and hence allows for heterogeneity among countries. It then asks how these endogenously determined groups differ from each other. The main analysis is based on time-series methods and therefore avoids some of the difficult issues resulting from the open-endedness of growth theories.

Specifically, in order to account for the variety of growth patterns across countries, the growth process is assumed to follow a Markov switching model. In this approach there are different growth regimes, which correspond to different growth performances such as fast growth or stagnation and which are equal across countries. The growth process of a country depends on how the country changes between the growth regimes and how long it stays within a growth state.[2] The dynamics of the switching process are captured by the transition matrix, which describes the probability of moving from one state to another. Since countries differ in their growth patterns, there is no reason to believe that the transition matrices are equal for all countries. In order to capture the most important differences in the transition matrices across countries, the Markov switching framework is enriched by a clustering framework,[3] which renders it possible to group countries into clusters with similar growth processes, i. e. with (close to) equal transition matrices. Once these clusters are known they can be compared along different dimensions such as the availability of human capital or the quality of institutions. The dimensions along which the comparisons are drawn should reflect "growth fundamentals" rather than short-term events such as an unfavorable change in the

---

[1] These problems are discussed in several survey articles such as Brock and Durlauf (2001), Durlauf (2001) or Temple (1999).

[2] The terms growth regimes and growth states are used interchangeably in this paper.

[3] The terms classification and clustering are used interchangeably in this paper.

terms of trade. The latter event would be reflected by a switch between growth regimes whereas the former variables have longer lasting impacts on the dynamics of growth.

The Markov switching classification model is estimated for the growth rate of GDP per capita in purchasing power parity for 84 countries during the period from 1962 to 2002. Four growth regimes are identified: a stable growth regime, which is characterized by steady growth, a miracle growth regime characterized by sustained high growth rates, a stagnation regime, which is volatile and on average features zero to slightly negative growth rates and a crisis regime that captures extreme growth behavior. The countries are divided into three clusters: the first cluster comprises successful countries that are characterized by long periods of stable and miracle growth, the second cluster comprises moderately successful countries that switch between stagnating and growing steadily and the third cluster comprises countries commonly referred to growth failures. Accordingly, these countries spend most of their time either in stagnation or in crisis. A descriptive and quantitative analysis attempts to reveal the growth fundamentals responsible for these different growth dynamics. Successful countries have generally succeeded in building up trustworthy institutions. Moreover, they possessed reasonable amounts of human capital at the beginning of the sample period and they have been open for most of the time. The quality of institutions, however, is not useful to distinguish between moderately successful countries and growth failures. The main distinguishing feature in this case seem to be the availability of human capital and of advantageous geographic conditions. The importance of openness cannot be verified quantitatively. However, the descriptive analysis certainly points into that direction.

This paper is related to previous applications of clustering techniques and Markov switching models in the economic growth literature. The idea of modeling the growth process as a Markov-switching model has been put forward by Pritchett (2003) and taken to the data by Jerzmanowski (2006). Jerzmanowski (2006) estimates a Markov switching model with four growth regimes. He allows for different transition matrices between countries, but conditions them on an indicator of the quality of institutions. The clustering of countries in the empirical growth literature has been suggested as a means to tackle the issue of parameter heterogeneity. In a seminal contribution Durlauf and Johnson (1995) drew attention to the fact that the marginal impacts of growth correlates need not be homogeneous across countries and that indeed the hypothesis of homogeneity is rejected when taken to the data. In order to identify the country clusters, they employed a regression-tree analysis, which essentially groups countries according to endogenously determined thresholds with regard to predetermined variables. More recently, threshold regressions have been applied. The intuition of the clustering is the same, but contrary to the regression tree analysis the asymptotic distribution of the estimates is known (Hansen, 2000; Papageorgiou, 2002). The major drawback of both methods is the need to explicitly name and determine the threshold variable beforehand, a decision that to some extent predetermines the clusters. A more data-based

3

approach is the application of mixture analysis. In this modeling framework cluster membership is interpreted as a latent variable that is estimated at the same time as the rest of the parameters. No prior specification with regard to a threshold variable is required. Examples of the latter approach can be found in Bloom et al. (2003), Paap et al. (2005), Basturk et al. (2008) or Alfo et al. (2008) taking the "classical" estimation approach and in Ardic (2006) taking the Bayesian estimation approach. Other clustering methods such as projection pursuit (Desdoigts, 1999; Kourtellos, 2002) or the predictive density approach (Canova, 2004) have been proposed, but they have not yet been widely used.

The contribution of this paper to the literature is threefold: It applies the Markov switching classification method introduced in the computer science literature to an economic problem. Unlike other clustering approaches the suggested method is aimed at deriving similarities in the patterns of growth rates instead of similarities related to marginal effects. Compared to the empirical study by Jerzmanowski (2006), the a priori determination of factors potentially influencing the transition probability matrix is avoided. Apart from the added flexibility the estimated Markov model avoids any issues that might arise in Jerzmanowski's work due to the fact that the transition probabilities depend on the potentially endogenous institutional quality variable (Glaeser et al., 2004).[4] Finally, previous approaches emphasizing the instability of growth rates have concentrated on determining the factors that start episodes of high or low growth (Hausmann et al., 2005, 2006; Aizenman and Spiegel, 2007; Jong-A-Pin and de Haan, 2007; Jones and Olken, 2008). Yet, for instance Hausmann et al. (2005) remark that initiating a growth episode might well require different measures than sustaining it. The present study separates shocks that initiate growth episodes from structural variables that determine the general conduciveness to and sustainability of growth and by construction handles both problems simultaneously. The results offer important hints at control variables that should be included when studying growth accelerations.[5]

The remainder of the paper is organized as follows. In Section 2 the theoretical Markov switching growth model is introduced and its empirical implementation is discussed. In section 3 the Markov switching classification method used for the estimation is discussed before the results are presented in section 4. Section 5 addresses the question what underlying features can explain the derived country clusters before section 6 concludes.

---

[4] Endogenous regressors in a Markov switching model generally render the estimates inconsistent if the endogeneity is not properly accounted for (Kim, 2004a,b; Kim et al., 2008).

[5] A further study concerned with the sustainability of growth accelerations is Berg et al. (2008).

2 Framework of Analysis

2.1 An encompassing growth model

The idea to model growth as a Markov switching process was put forward by Pritchett (2003) following an enumeration of empirical facts that an encompassing growth model should be capable to explain, but previous growth models cannot. In Pritchett's view an encompassing growth model has to include mechanisms that can generate the divergence observed in absolute and relative income levels since 1870. Moreover, it has to be capable of explaining the sustained steady growth rates observed in industrial countries, the extremely rapid growth rates of newly industrializing countries, but also the zero growth experiences of many poor countries. In addition to that, it has to account for the instability of growth rates especially in developing countries as well as the empirically observed parameter instability. Pritchett argues that there is no unified growth theory that can address all of these issues simultaneously. Hence, one way to deal with this difficulty is to define different growth regimes and assume that within each regime a well defined growth model governs the growth process. In order to account for the development process of economies and the observed instability of growth rates, a country has to be able to switch between the growth regimes. It follows that the growth theories prevailing in each regime have to be supplemented by explanations as to why and when a country will switch between them. This aspect of the model is captured in the accompanying transition probability matrix of the Markov switching model.

Pritchett (2003) suggests a growth model comprising six growth regimes: The first one explains steady state growth of high income industrial countries. An appropriate model for this growth state could be the Solow model or endogenous growth theories. The second regime deals with countries that find themselves in an underdevelopment trap,[6] i. e. that experience zero growth at very low income levels. The third regime captures countries that grow continuously with a similar growth rate as industrial countries but at lower levels of income and are therefore neither converging nor diverging in relative terms. A fourth regime has to explain the rapid growth experiences such as those observed in newly industrial countries and a fifth state has to explain episodes of growth implosions such as the large collapses of output in "transition" countries at the beginning of the 1990s. Finally, the sixth regime has to explain zero growth at medium income levels. With regard to the transition probabilities Pritchett suggests that they depend on initial conditions, history, policies and institutions as well as previous states. Since the states are defined in terms of both growth rates and income levels, the transition probabilities have to be asymmetric. For example, it is logically impossible to change from being a country in state two (subsistence level) to being a country in state one (leading economy).

---

[6] The wording underdevelopment trap is preferred as opposed to poverty trap, as the latter could apply both to macroeconomic and a microeconomic aspects. This paper, however, focuses on the macroeconomic aspects only. Cf. Berthelemy (2006).

2.2 Empirical implementation

Empirically, it is too demanding to apply Pritchett's (2003) suggestions directly to the data. In order to estimate a variant of the model, Jerzmanowski (2006) simplifies the model as follows. The states are defined solely in terms of growth rates and the growth process in each state is represented by a simple AR(1) process, thereby avoiding the need to explicitly connect states with growth theories. The transition between states is governed by a Markov chain of order one, thus accounting for the dependence of transition on the previous state. The transition probabilities of the Markov chain are defined as country-specific and are estimated conditional upon the institutional quality as measured by the index of government anti-diversion policies averaged over the years 1986-1995.[7] Formally, the growth rate in each state is given by

$$y_{tk} = \alpha_{s_t} + \beta_{s_t} y_{t-1k} + \varepsilon_{tk}^{s_t}, \text{ with } \varepsilon_{tk}^{s_t} \sim \text{i.i.d. } N(0, \sigma_{s_t}^2), \tag{1}$$

where y_{tk} is the growth rate of country k in period t and where s_t indicates the growth state that the country finds itself in. The growth rate in each period is influenced by a regime-specific random shock ε.[8] The evolution of the state variable is governed by the following first-order Markov chain, where

$$p_{ij} = P(s_t = j | s_{t-1} = i, s_{t-2} = i_2, ...) \quad = \quad P(s_t = j | s_{t-1} = i)$$

denotes the probability of changing from state i in the previous period to state j in the current period.[9] The first-order Markov chain assumption ensures that only last period's state, but not the entire history of states influences the transition probabilities. Finally, the transition probabilities are made country-specific by making the transition probabilities depend on a vector of exogenous country-specific time-independent variables z_k, i. e.

$$p_{ij}(z_k) \equiv P_k(s_t = j | s_{t-1} = i). \tag{2}$$

---

[7] A more comprehensive definiton is given in section 5.

[8] The chosen formulation abstracts from common shocks which influence several countries at the same time. Some information about the appropriateness of this assumption can be obtained from studies dealing with international business cycle dynamics. Kose et al. (2003) show that common factors are important for developed countries, whereas they are not for developing countries. In a follow-up study Kose (2005) show that regional factors are becoming more important for emerging market economies, but this is only a recent development. Stock and Watson (2005) note that common factors for business cycles dynamics in industrialized countries have become less important over time. It therefore appears to be a defendable simplification to abstract from common shocks, which at best are important for a minor number of countries in the chosen framework. We will retain this simplification throughout the paper.

[9] $\sum_{\forall j} p_{ij} = 1$ holds. Cf. section 3.

z_k includes a constant and the institutional quality measure. $p_{ij}(z_k)$ is specified in logit form.

Jerzmanowski (2006) estimates the Markov switching model over the years 1962-1994 for 89 countries and obtains four growth states (see table 2.2): state one reflects stable growth with steady state growth rates around 2 %, state two is a stagnation state with zero steady state growth, state three is a crises state with negative growth rates and state four is a miracle growth state with a steady state growth rate of 6 %. Compared to Pritchett's (2003) original suggestion the former states one and three are summarized in the new state one and the former states two and six are summarized in state two. Both changes can be attributed to the fact that income levels are no longer a defining element of the states. Focusing on the influence of institutional quality on the transition probabilities, Jerzmanowski finds that countries with a high quality of institutions will find themselves in the stable growth state with a high probability, whereas countries with a low quality of institutions are more likely to find themselves in the stagnation state. The miracle growth state is most often visited by countries with medium quality institutions.

The results of the estimation are very appealing. Apart from the good interpretablity of the states, Jerzmanowski (2006) shows that the estimated regime changes often coincide with actual important political or economical events and that countries spend reasonable amounts of time in each of the regimes. One shortcoming of the estimation procedure, however, concerns the conditioning of the transition probability on the quality of institutions. Pritchett (2003) suggests a whole set of variables that are likely to influence the transition probabilities. It therefore seems problematic to pick just one variable and omit all the others. An additional problem may arise from the fact that the quality of institutions is measured at the end of the sample period. Since the causality between growth and the quality of institutions is not yet settled (Glaeser et al., 2004), it is conceivable that economic performance influences the quality of institutions. If this were the case, the estimates would be inconsistent (Kim et al., 2008).[10] The usual approaches to include additional control variables in the transition probability matrix or to mitigate the potential simultaneity problem by instrumental variables are infeasible for the problem at hand because both procedures would lead to even higher levels of non-linearity. Therefore, we suggest using a clustering approach in order to shed more light on the factors determining the transition probabilities. Since different transition probabilities are derived from the data itself without conditioning them on additional variables, the danger of an endogeneity problem is avoided.

The model design in this paper essentially equals that of Jerzmanowski (2006). The growth

---

[10] Kim et al. (2008) suggests a method to test for the endogeneity of the conditioning variable, but this would require a more general solution of the highly nonlinear problem, which at the moment appears infeasible.

Table 1: Estimation results from Jerzmanowski (2006)

	Constant	AR Coefficient	Standard Deviation	Steady State Growth Rate
State 1	1.32*	0.3761*	2.11	2.12
State 2	0.10	0.1799*	4.56	0.12
State 3	−1.01*	−0.0045	13.16	1.00
State 4	5.36*	0.1417*	2.71	6.25

Except for the AR coefficients all numbers are percentages. The stars denote the significance of the estimate at the 5 % level.

rates in each growth regime are modelled as AR(1) processes, i. e.

$$y_{tk} = \alpha_{s_t} + \beta_{s_t} y_{t-1k} + \varepsilon_{tk}^{s_t}, \text{ with } \varepsilon_{tk}^{s_t} \sim \text{i.i.d. } N(0, \sigma_{s_t}^2). \tag{3}$$

Each growth regime continues to be characterized by regime specific coefficients and a regime specific variance. The crucial difference concerns the transition probabilities. While the evolution of the state variable continues to be governed by a first-order Markov chain, the transition probabilities are no longer made country-specific. Instead, it is assumed that the data is generated by m different transition probability matrices. Countries the data of which has been generated by the same transition probability matrix are grouped together in a cluster. Hence, the transition probabilities are given conditional on belonging to a certain cluster. Let C_m denote membership in cluster m. The transition probabilities in the clusters are defined as

$$p_{ij}(C_m) \equiv P_{C_m}(s_t = j | s_{t-1} = i). \tag{4}$$

Cluster membership is endogenously determined in the estimation process. Compared to Jerzmanowski (2006) this model specification is less restrictive in that countries with the same quality of institutions may feature different transition probabilities. The specification is more restrictive in that not every country is allowed to have a unique transition probability matrix. Nevertheless, an analysis of the resulting clusters seems to be a promising way to learn more about the forces driving the transition dynamics in such a growth model.

3   Methodology

The proposed model interprets the Markov switching model with constant transition probabilities in a panel context and extends it by a clustering mechanism. In order to facilitate the presentation of the estimation procedure, we will first briefly summarize the estimation strategy for simple Markov switching models before introducing the idea of the clustering mechanism. After that, the details of the parameter estimation are presented before some issues regarding the methodology are discussed.

3.1   The Basic Markov Switching Model

Consider the following Q-state Markov switching model:

$$y_t \;=\; x_t' \beta_{s_t} + \varepsilon_t^{s_t}, \qquad t = 1, ..., T. \tag{5}$$

The evolution of the dependent variable y depends on the observed exogenous variables x,[11] which may include autoregressive terms, *and* the contemporaneous state of the model $s_t \in \{1, ..., Q\}$ because the marginal effects of the exogenous variables equalling β_{s_t} depend on the state. Moreover, the residuals are state-dependent. Within each regime they follow a normal distribution $N(0, \sigma_{s_t}^2)$ with a regime-specific variance.

The transition probabilities are summarized in matrix $\mathbf{P}$

$$\mathbf{P} = \left[\begin{array}{cccc} p_{11} & p_{21} & \cdots & p_{Q1} \\ p_{12} & p_{22} & \ddots & p_{Q2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1Q} & p_{2Q} & \cdots & p_{QQ} \end{array} \right], \tag{6}$$

where $p_{ij} = P(s_t = j | s_{t-1} = i)$ and $\sum_j p_{ij} = 1, \forall j$ hold. $\mathbf{P}$ is assumed to be ergodic, i. e. the eigenvalues of $\mathbf{P}$ lie inside the unit circle with the exception of one eigenvalue equal to one (Hamilton, 1994, chap. 22).

If the states of the process were known with certainty, the parameters of the Markov switching model could easily be estimated. One could simply run an OLS regression in the form of equation (5) for each state separately. The transition probabilities in (6) could be derived by counting the number of times state i is followed by state j divided by the number of observations. Alternatively, the complete-data log-likelihood function of the Markov switching model could be formulated and

---

[11] x_t may be a $(v \times 1)$ vector of exogenous variables, $v \geqslant 1$. Accordingly, β_{s_t} may be a parameter or a vector of parameters.

maximized, whereby complete refers to the fact that the latent state variables are known. Let $I(s_t = j)$ denote an indicator function that takes on the value one if state j prevails in period t and zero otherwise. Then the complete-data log-likelihood function can be shown to be given by[12]

$$
\begin{aligned}
\ell_c(\mathcal{Y}_T, \mathcal{S}_T | \mathcal{X}_T, \theta) = {} & \sum_{j=1}^{Q} I(s_1 = j) \left(\log f(y_1 | s_1 = j, \psi_0, x_1, \theta) + \log \rho_j \right) + \\
& \sum_{t=2}^{\mathbb{T}} \left\{ \sum_{j=1}^{Q} I(s_t = j) \left(\log f(y_t | s_t = j, \psi_{t-1}, x_t, \theta) \right) + \right. \\
& \left. \sum_{i=1}^{Q} \sum_{j=1}^{Q} I(s_t = j | s_t = i) \log p_{ij} \right\}.
\end{aligned}
\tag{7}
$$

$f(\cdot)$ denotes the conditional density of y_t. ρ_j is a shortcut for $P(s_1 = j | \psi_0, \theta)$. ψ_t denotes the information up to period t, which consists of all observed dependent and independent variables up to period t denoted by $\mathcal{Y}_t$ and $\mathcal{X}_t$. The history of states up to t is denoted $\mathcal{S}_t$. The parameters of the model are summarized in θ. Hence, θ consists of $(\beta_1, \ldots, \beta_Q)$, $(\sigma_1^2, \ldots, \sigma_Q^2)$ and the transition probabilities. The ergodic distribution of $\mathbf{P}$ is used to predict the initial probabilities ρ_j, $\forall j = 1, \ldots, Q$. Maximization of (7) is straightforward because the first derivatives with respect to the parameters characterizing the AR-process and the variances are independent of any expressions involving p_{ij} $(\forall i, j)$ and vice versa.

Unfortunately, in reality the prevailing states are not directly observable and have to be inferred from the observed values of the exogenous and endogenous variables, which makes the use of more sophisticated estimation procedures necessary. One popular estimation procedure in this context is the expectation-maximization algorithm, in short the EM algorithm, suggested by Dempster et al. (1977).[13] The basic idea of the EM algorithm is to first replace the state indicator functions with the best guess for s_t, the so called expectation step, and to estimate the parameters θ conditional on this best guess, the so called maximization step. Any best guess of s_t necessarily depends on θ. Therefore, the estimation and maximization steps are iterated until convergence of θ is attained. The limit of the iterations corresponds to the maximum likelihood estimator, but is much easier to obtain because the dominant source of nonlinearity involved in the joint estimation of s_t and θ is reduced by means of the expectation step.

Hamilton (1990) shows that the best guess for s_t in the expectation step is obtained if the indicator functions in the complete-data log-likelihood function are replaced by the so called smoothed

---

[12] A more detailed derivation of the complete-data log likelihood function can be found in appendix A.

[13] Other popular estimation procedures are the numerical maximization of the incomplete-data log-likelihood function by summing over all possible state histories (Hamilton, 1989) or Bayesian estimation procedures (Fruehwirth-Schnatter, 2006).

state probabilities. The new log-likelihood function is the expected complete-data log-likelihood function, which will also be referred to as the incomplete-data log-likelihood function in the following. The smoothed state probabilities can be calculated using special filtering and smoothing algorithms, which Krolzig (1997) labels the Baum-Lindgren-Hamilton-Kim (BLHK) filter. The filter first determines the probability that the tth observation has been generated by regime i conditional on a parameter value θ and conditional on the information up to period t, i. e. the filtered probability $P(s_t = i|\psi_{t-1}, y_t, x_t, \theta)$. Based on the filtered probability it is possible to form a forecast of how likely the process is to be in regime j in period $t+1$ given the observations up to period t, formally $P(s_{t+1} = j|\psi_{t-1}, y_t, x_t, \theta)$. Suppose that the Q conditional densities $f(y_t|s_t = i, \psi_{t-1}, x_t, \theta)$ are collected in the $(Q \times 1)$ matrix η_t and that the Q filtered and predicted state probabilities $P(s_t = i|\psi_{t-1}, y_t, x_t, \theta)$ and $P(s_{t+1} = j|\psi_{t-1}, y_t, x_t, \theta)$ are collected in the $(Q \times 1)$ matrices $\xi_{t|t}$ and $\xi_{t+1|t}$, respectively. Then the filtered and predicted state probabilities can be obtained by iterating forwards on the following two equations:[14]

$$\hat{\xi}_{t|t} = \frac{\hat{\xi}_{t|t-1} \odot \eta_t}{1'(\hat{\xi}_{t|t-1} \odot \eta_t)} \tag{8}$$

$$\hat{\xi}_{t+1|t} = \mathbf{P}\dot{\xi}_{t|t}. \tag{9}$$

In order to start the filter, $\hat{\xi}_{1|0}$ is assumed to equal the ergodic distribution implied by the transition probabilities.[15]

The filtered and predicted probabilities are the best state inferences and forecasts available if in each period the information up to that period is used. Based on them the smoothed state probabilities can be obtained. Unlike the filtered state probabilities the smoothed state probabilities always use all the available information in the sample, i. e. the entire time series, to infer the most likely state in period t. The smoothed state probabilities $P(s_t = i|\psi_T, \theta)$ are derived by iterating backwards on equation (10), where the Q smoothed probabilities of period t are collected in the $(Q \times 1)$ matrix $\xi_{t|T}$.

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} \odot \{\mathbf{P}' \cdot [\hat{\xi}_{t+1|T} \oslash \hat{\xi}_{t+1|t}]\} \tag{10}$$

The backwards recursion is started by using $\hat{\xi}_{T|T}$ obtained from the filtered state probabilities.[16]

3.2 Clustering using Classification Maximum Likelihood

The basic Markov switching model has only been formulated for a single time series. However, it is straightforward to extend the model to panel data if the countries are assumed to share the

---

[14] The circle symbols such as $\odot$ indicate element-by-element operations. 1 represents a $(Q \times 1)$ vector of ones.

[15] Cf. Hamilton (1994, p. 693) for further implementation possibilities.

[16] This section draws heavily on Hamilton (1994, chap. 22), Krolzig (1997, chap. 5 and 6), and Kim and Nelson (1999, chap. 4). The presentation of the complete-data log-likelihood function follows Diebold et al. (1994).

same parameters and if growth in each country is independent of the events in other countries. In this case the EM algorithm remains unchanged apart from the fact that the BLHK-filter has to be started separately for each country.[17]

In order to account for possibly differing transition probabilities across countries, the panel Markov switching model is further extended by allowing countries to belong to different clusters. While all clusters share the same equation (5), they feature unique transition probabilities. Hence, each cluster implies a distinctive data-generating process. As in the previous section we will first establish the complete-data log-likelihood function for this revised problem before proceeding to the approach used in the presence of the latent variables. In order to facilitate the exposition, the complete time-series for country k will be denoted as $\mathcal{O}^k$.

Suppose that $m = 1, \ldots, M$ different clusters are allowed for. These clusters share the same coefficients and the same variances, but differ with respect to $\mathbf{P}$. Collect the parameters of each cluster in θ_m and summarize all θ_m in the parameter vector θ. If the states at each point in time are known and if it is also known that $\mathcal{O}^k$ is generated by cluster m, the complete-data log-likelihood function for $\mathcal{O}^k$ equals equation (7) with θ being replaced by the appropriate θ_m. In the following $\ell_c(\mathcal{Y}_{kT}, \mathcal{S}_{kT} | \mathcal{X}_{kT}, \theta_m)$ is abbreviated by $\ell_{ck}(\mathcal{O}^k | \theta_m)$. As in the previous section a further indicator function $C_m(\mathcal{O}^k)$ is introduced in order to set up the complete-data log-likelihood function for the panel as a whole. $\mathcal{C}_m(\mathcal{O}^k)$ takes on the value one if $\mathcal{O}^k$ is generated by cluster m and zero otherwise. The complete-data log-likelihood function extended for a panel and including cluster membership information is given by[18]

$$\mathcal{L}_c(\mathcal{Y}_T, \mathcal{S}_T, \mathcal{C} | \mathcal{X}_T, \theta) = \sum_{k=1}^{K} \sum_{m=1}^{M} \mathcal{C}_m(\mathcal{O}^k) \ell_{ck}(\mathcal{O}^k | \theta_m). \tag{11}$$

Since neither the states nor the cluster memberships are known in reality, they have to be inferred from the observed data. The true states are approximated by the smoothed state probabilities. Each time series is assumed to have been generated by the cluster for which the expected complete-data log-likelihood function conditional on the states is maximized. All the time-series that have been generated by cluster m share the same transition probabilities. The EM algorithm from the previous section is augmented by a classification step. Given a current value for θ the smoothed state probabilities for each observation in each cluster are calculated (expectation step).

[17] Panel Markov switching models are not very common in the economics literature. Some examples are Asea and Blomberg (1998) and Chen (007a,b). These authors estimate panel Markov switching models with fixed effects and they also retain the assumption of no correlation across time series. This assumption is certainly worth relaxing in the future, but the currently limited data on growth does not allow for such an approach here. See also footnote 8.

[18] Owing to the panel context it now holds that $\mathcal{Y}_T = \{\mathcal{Y}_{1T}, \ldots, \mathcal{Y}_{KT}\}$, with K being the total number of countries in the analysis. The same applies to $\mathcal{S}_T$ and $\mathcal{X}_T$. $\mathcal{C}$ summarizes the cluster membership of countries.

Next, the expected complete-data log-likelihood values for each time series in each cluster are derived. Each time series is allocated towards the cluster which exhibits the highest log-likelihood value (classification step). This method of classification ensures that $\mathcal{L}_c(\mathcal{Y}_T, \mathcal{S}_T, \mathcal{C}|\mathcal{X}_T, \theta)$ never falls.[19] Finally, the parameters are reestimated conditional on the smoothed state probabilities and the cluster classification. These steps are iterated until convergence is achieved. In this case, convergence means that both the cluster memberships do not change any longer and that a convergence criterion for θ is met. Usually, cluster memberships only change in the first iterations and remain constant afterwards.[20]

3.3 Parameter Estimation

In this section the formulas needed for the estimation of the model are derived. The estimation is carried out by implementing the algorithm in Matlab. For convenience we briefly repeat the model from section 2. The growth model consists of $m = 1, \ldots, M$ Markov switching clusters with Q states each. Within one regime the growth rate evolves according to an AR(1) process with state-specific coefficients, which are required to be equal across all M clusters, and regime-specific error terms. Formally:

$$y_{tk} = \alpha_{s_t} + \beta_{s_t} y_{t-1k} + \varepsilon_{tk}^{s_t}, \text{ with } \varepsilon_{tk} \sim \text{i.i.d. } N(0, \sigma_{s_t}^2) \tag{12}$$

The M Markov switching clusters differ with respect to their transition probabilities. Each cluster features its own transition probability matrix $\mathbf{P}^m$, whereby

$$\mathbf{P}^m = \begin{pmatrix} p_{11}^m & p_{21}^m & \cdots & p_{Q1}^m \\ p_{12}^m & p_{22}^m & \cdots & p_{Q2}^m \\ \vdots & \vdots & \ddots & \vdots \\ p_{1Q}^m & p_{2Q}^m & \cdots & p_{QQ}^m \end{pmatrix}. \tag{13}$$

The interpretation of p_{ij}^m is equal to that of p_{ij}, but refers to cluster m.

Using the same indicator functions as in the previous section and making use of the normal

---

[19] Cf. Appendix B for a more detailed explanation.
[20] The clustering method described here has previously been implemented for instance by Alon et al. (2003) and Knab (2000).

distribution assumption, the complete-data log-likelihood function can be written as

$$\mathcal{L}(\mathcal{Y}_t, \mathcal{S}_T, \mathcal{C}|\mathcal{X}_t, \theta) = \sum_{k=1}^{K} \left\{ \sum_{m=1}^{M} \mathcal{C}_m(\mathcal{O}^k) \left\{ \sum_{j=1}^{Q} I(s_t = j) \right. \right.$$

$$\left(-\frac{1}{2} \log(2\pi) - \frac{1}{2} \sigma_j^2 - \frac{1}{2} \frac{(y_{1k} - \alpha_j - \beta_j y_{0k})^2}{\sigma_j^2} + \log \rho_j \right)$$

$$+ \sum_{t=2}^{T} \left[\sum_{j=1}^{Q} I(s_t = j) \left(-\frac{1}{2} \log(2\pi) - \frac{1}{2} \sigma_j^2 - \frac{1}{2} \frac{(y_{tk} - \alpha_j - \beta_j y_{t-1k})^2}{\sigma_j^2} \right) \right. \tag{14}$$

$$\left. \left. \left. + \sum_{i=1}^{Q} \sum_{j=1}^{Q} I(s_t = j, s_{t-1} = i) \log p_{ij}^m \right] \right\} \right\}.$$

In the following one round of the iteration process is described. All formulas can be verified by taking the first derivatives of the expected complete-data log-likelihood function. At the beginning of each iteration an estimate of θ and a classification of the time-series into clusters is available. We do not introduce separate notation to indicate that the coefficients are estimated nor do we introduce further superscripts to refer to the round of iteration that the estimates result from. These issues are clarified by the accompanying notes. However, the labels $\tilde{\ell}_{ck}(\mathcal{O}^k|\theta_m)$ and $\tilde{\mathcal{C}}_m(\mathcal{O}^k)$ are introduced and denote the incomplete-data log-likelihood function of country k in cluster m and the derived country classification indicators, respectively.

1. Given the current parameter estimate of θ compute $P_k^m(s_t = j|\psi_T, \theta_m) \ \forall j, k, m, t$, i. e. the smoothed state probabilities at every point in time for every country k in every conceivable cluster m using equations (8), (9) and (10).

2. Given the updated smoothed state probabilities and the current estimate of θ compute the new values for the expected complete-data log-likelihood functions $\tilde{\ell}_{ck}(\mathcal{O}^k|\theta_m)$ for every country k in every conceivable model m. If there is a $\tilde{\ell}_{ck}(\mathcal{O}^k|\theta_m)$ greater than $\tilde{\ell}_{ck}(\mathcal{O}^k|\theta_m)$ of the cluster that country k is currently assigned to, reassign country k to the cluster maximizing $\tilde{\ell}_{ck}(\mathcal{O}^k|\theta_m)$ over all m. Else do not change the cluster assignment.

3. Given the updated smoothed state probabilities and the updated country classification derive the new transition probabilities as

$$p_{ij}^m = \frac{\sum_{k=1}^{K} \sum_{t=2}^{T} \tilde{\mathcal{C}}_m(\mathcal{O}^k) P_k^m(s_t = j, s_{t-1} = i|\psi_T, \theta_m)}{\sum_{k=1}^{K} \sum_{t=2}^{T} \tilde{\mathcal{C}}_m(\mathcal{O}^k) P_k^m(s_{t-1} = i|\psi_T, \theta_m)}. \tag{15}$$

The value of ρ_j^m corresponds to the probability of state j in the ergodic distribution associated with $\mathbf{P}^m$.

4. Summarize the coefficients α_j and β_j for each state in $\kappa_j = (\alpha_j, \beta_j)'$. Collect the explanatory variables in $z_{tk}' = [1 \ y_{t-1k}]$. Given the updated smoothed probabilities the updated parameter

estimate κ_j is obtained as

$$\kappa_j = \left[\sum_{k=1}^{K}\sum_{t=1}^{T}\tilde{z}_{tk}(j)\tilde{z}'_{tk}(j)\right]^{-1}\left[\sum_{k=1}^{K}\sum_{t=1}^{T}\tilde{z}_{tk}(j)\tilde{y}_{tk}(j)\right] \qquad (16)$$

where

$$\tilde{z}_{tk}(j) = z_{tk} \odot \sqrt{\sum_{m=1}^{M}\tilde{\mathcal{C}}_m(\mathcal{O}^k)P_k^m(s_t = j|\psi_T, \theta^m)} \text{ and} \qquad (17)$$

$$\tilde{y}_{tk}(j) = y_{tk} \cdot \sqrt{\sum_{m=1}^{M}\tilde{\mathcal{C}}_m(\mathcal{O}^k)P_k^m(s_t = j|\psi_T, \theta^m)}. \qquad (18)$$

In practise, a separate OLS regression on the observations weighted by the square root of the smoothed probabilities is carried out for each state.

5. Given the updated smoothed probabilities and the updated parameter estimates, the updated regime-specific variances are obtained as

$$\sigma_j^2 = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{m=1}^{M}\tilde{C}_m(\mathcal{O}^k)(y_{tk} - \alpha_j - \beta_j y_{t-1k})^2 P_k^m(s_t = j|\psi_T, \theta^m)}{\sum_{k=1}^{K}\sum_{t=1}^{T}\sum_{m=1}^{M}\tilde{C}_m(\mathcal{O}^k)P_k^m(s_t = j|\psi_T, \theta^m)} \qquad (19)$$

6. Check whether convergence is achieved. Convergence requires a stable country classification. Moreover, the relative change in the expected log-likelihood function must not exceed 10^{-5}. If no convergence is achieved, restart the algorithm at step 1.

3.4 Some Remarks

In this section we want to address some critical issues concerning the EM algorithm and the classification approach that have been discussed in the literature. The focus will be on the choice of initial values for starting the algorithm, on model selection and on the merits of the proposed method in the present context.

The sensitivity of the EM algorithm with respect to the starting values of the iterative process is a well known weakness (Karlis and Xekalaki, 2003; McLachlan and Krishnan, 1997, chap. 4; Biernacki et al., 2003). The initial values determine both the speed of convergence and the ability to locate the global maximum of the problem. For instance, depending on the starting values the algorithm might get trapped in a very flat area of the log-likelihood function and might falsely assume that it has reached a local maximum. Moreover, in the case of multiple local maxima the initial values determine which of the maxima is attained. Clearly, the choice of appropriate

starting values is important. The literature offers three ways to proceed:[21] One possibility is to start the algorithm simply from different, possibly random, initial values. Alternatively, starting values might be obtained either from prior theoretical and/or empirical knowledge or by using a different easily implementable estimation method beforehand such as k-means clustering. Finally, the expectation step in the EM algorithm might be replaced by a stochastic simulation step, which reduces the tendency of the algorithm to get trapped in a local maximum or to find a spurious solution.[22] Karlis and Xekalaki (2003) and Biernacki et al. (2003) recommend to use an adaptation of the first method in practise. Instead of iterating the EM algorithm for each starting value until convergence, they suggest running the EM algorithm from a large number of initial values, but only for a small number of iterations. The solutions which maximize the expected loglikelihood function should then be selected for complete runs of the EM algorithm. This approach will be applied in the following. The EM algorithm will be started using 1000 random parameter values and computed for 25 iterations. Of the 1000 random starting values the EM algorithm will be iterated until convergence only for those 10% yielding the highest expected log-likelihood value after 25 iterations.

Up to now we have always assumed that the number of states Q and the number of clusters M are known. With regard to the number of states, the encompassing growth model and the empirical implementation by Jerzmanowski (2006) suggest that a total number of four states is appropriate if the growth regimes are only defined via the growth rate disregarding the prevailing income level. However, the required number of clusters is unknown so that some model selection strategy is called for. Formal tests in this environment are difficult to implement because model selection involves inference for an overfitted model, where the true number of clusters is less than the number of clusters in the fitted model. In such a situation the traditional LR-test can no longer be applied because the parameters of the overfitted, i. e. the alternative model, are not identified under the null hypothesis. Therefore, even if the additional parameters are estimated to be significant, it is still possible that this significance is simply due to sampling variation (Hansen, 1992; Garcia, 1998).[23] Given the difficulties in implementing formal testing procedures, the most popular method for inferring the number of clusters is the use of information criteria. For the clustering of temporal data in a Markov switching model Li and Biswas (2000) suggest using the Bayesian information criterion (BIC). However, whereas traditionally the number of estimated parameters

---

[21] Even though the literature with respect to the starting values deals predominantly with finite mixture models, it is important to consider this problem in the estimation of Markov switching models, too, because they have been shown to be at least as sensitive to starting values as mixture models (Dunmur and Titterington, 1998).

[22] Several methods are shown in Biernacki et al. (2003), and McLachlan and Krishnan (1997).

[23] The derivation of the required number of clusters resembles that of deriving the correct number of states in simple Markov models. For this case, modified LR tests have been proposed, but their implementation is quite complicated (cf. Krolzig (1997, Section 7.5) and the references therein). Therefore, the number of regimes is usually derived using information criteria. Recently, Smith et al. (2006) have suggested an information criterion, which has been derived specifically to establish the number of required states in a Markov switching model with constant transition probabilities.

are penalized by the logarithm of the total number of observations, in the clustering context only the total number of cross-sections should be used. Alon et al. (2003) derive a very similar criterion based on the minimum description length. This criterion, which will be termed the modified BIC in the following differs from the previous BIC in that the penalty term is applied on the number of estimated parameters and the number of clusters. Often, the use of the consistent Akaike information criterion (CAIC) is suggested in the context of latent class models (Basturk et al., 2008; Jedidi et al., 1997). In the next section all three information criteria will be used to determine the required number of clusters.

A final remark is necessary with regard to the type of clustering chosen. In this paper a hard clustering approach is used, i. e. each time-series belongs to one and only one cluster at a given time. An alternative would be soft clustering, where cluster membership is represented probabilistically. An example of the latter approach are finite mixture models. In the absence of panel data it is well established that soft clustering is preferable to hard clustering. By construction, the probabilistic assignment to clusters allows an assessment of the confidence of the cluster assignments. More importantly, in the absence of panel data hard clustering has been shown to lead to reasonable clusters, but inconsistent parameter estimates (Celeux and Govaert, 1993; Bryant, 1991; McLachlan, 1982). Soft clustering in the context of Markov switching models is possible (Butler, 2003; Alon et al., 2003; Wichern, 2001; Cadez and Heckermann, 2003), but computationally very demanding and rarely used. In the context of the proposed model a further difficulty would arise: The Markov switching clusters in this paper differ only with respect to the transition probabilities, but not with respect to the state coefficients. This implies that by construction the differences in the incomplete/complete-data log-likelihood functions tend to be small so that the traditional smoothed model probabilities, i. e. the probability that given the parameters the data has been generated by cluster m, are too close to each other to allow for a soft clustering mechanism to be well defined. Despite opting for the hard clustering approach, our model does not suffer from the inconsistency problem pointed out in the hard clustering of mixture models. Since panel data is available the cluster assignment is consistent for large enough time-series. If the cluster assignment is consistent, so are the parameter estimates.[24] However, since we assign each time-series deterministically to one cluster only, the drawback of not being able to assess the accuracy of the cluster assignment remains.

The previous paragraph leads to the question to what extent the clustering results are trustworthy. Whereas it is not possible to evaluate the accuracy of the assignment of a single country separately, the use of simulation methods enable us to analyze how well on average the model

---

[24] For Markov switching parameter estimates it is typically assumed that they are $\sqrt{T}$-consistent and that the usual likelihood based methods of inference can be applied (Krolzig (1997, chap. 6.6 and 7); Psaradakis and Sola (1998)).

works in the present context. To that end, the parameter estimates for each number of clusters have been used to simulate time-series[25] and the percentage of time-series correctly classified has been calculated. Each simulation consists of 30 time-series. The results are presented in table 2.[26]

Table 2: Share of Time-Series Correctly Classified in Simulations

	M = 2	M = 3	M = 4
T = 41	0.9528	0.8647	0.8122
T = 50	0.9694	0.8933	0.8540
T = 100	0.9968	0.9798	0.9583

This table reports the share of time-series correctly classified if parameters very similar to the estimation results are used. Each simulated time-series has been estimated ten times and the best estimation results has been selected in order to generate the table.

As expected, the classification works better, the more observations are available. The classification also deteriorates with the number of available clusters. Whereas the first observation follows from the consistency of the assignment for large T, the second observation follows from the fact that the transition matrices tend to be more dissimilar the smaller the number of clusters. Hence, the assignment to the clusters is carried out with less uncertainty. The number of observations per time-series that is available in this paper is sufficient to obtain reasonable results. Even if four clusters are chosen, on average more than 80% of the time-series are correctly classified. Hence, while it can never be ruled out that a particular time-series is misclassified, overall the results are informative. Unlike the number of clusters, the number of growth regimes is not a serious restriction for the accuracy of the classification: since all countries visit the same four growth regimes, there are enough observations to identify them.

---

[25] Slight adjustments of the transition probability matrix have been made if a rounded entry equalled 0.000

[26] The estimation of each simulated time-series has been repeated ten times and the best results has been used in the table. It is conceivable that additional repetitions could have improved the classification results even further, so that the reported percentages should be seen as a lower bound for the success of the method.

4 Empirical Results

4.1 Parameter Estimates

In this section we take the model to the data. The data set consists of the annual growth rates of real GDP per capita (measured in purchasing power parity) of 84 countries for the period starting in 1962 and ending in 2002. The set of countries comprises 27 African, 19 Asian, 15 European, 11 North-American, 10 South-American and 2 Oceanian countries. The data is taken from the Penn World Tables, Version 6.2.[27] The Markov switching model is estimated taking the number of four different growth regimes as given. The number of clusters is varied between one and four. The model is estimated from a multitude of starting values as explained in section 3.4. For each solution Jarque-Bera tests on the regime-conditional error distribution in every cluster and for the model as a whole are conducted (Campbell, 2002). For all considered cluster-regime combinations the estimation results maximizing the incomplete-data log-likelihood function and simultaneously satisfying the normality assumption on the five percent significance level within each cluster and for the model as a whole are reported.

The first step is to determine the number of required clusters. For this purpose, the model is estimated using four states for $M = 1, \ldots, 4$ clusters. The resulting information criteria are reported in table 3. The first row contains the BIC as suggested by Li and Biswas (2000), the second one the modified BIC as suggested by Alon et al. (2003). The third row contains the consistent AIC, which like the other two information criteria is calculated using the number of cross-sections and not the total number of observations in the penalty term. The information criteria are formulated such that the model with the smallest value should be chosen. All information criteria consistently reject a model without different clusters in favor of a model containing three clusters.

Table 4 shows the estimated parameters of the autoregressive processes for each of the four states. The first column reports the estimated constants, the second one the estimated autoregressive coefficients and the third one the regime-specific standard errors. The fourth column indicates the implied steady state growth rates. The results are largely comparable to those obtained by Jerzmanowski (2006). The first reported state corresponds to Jerzmanowski's crisis regime. It is characterized by a negative constant, a negative autoregressive coefficient and a negative steady state growth rate. In fact, naming this state the crisis regime is somewhat misleading: due to the large standard error of this regime it not only captures episodes of extremely negative growth rates, but also episode of extremely positive growth rates. In our view, the most outstanding feature of

---

[27] The variable $RGDPCH$ is used. The growth rates are constructed by taking the first difference of the log levels. The countries were chosen on the basis that GDP per capita is available without interruptions from 1960 - 2002. Only countries that had a population of more than one million for at least half of the sample period were considered. Moreover, newly added countries to PWT 6.2 were disregarded because they suffer from implausibly high income estimates in the past as the appendix to the Penn World Tables points out.

Table 3: Information Criteria for Different Numbers of Clusters

	M = 1	M = 2	M = 3	M = 4
BIC	5.6370	5.5777	5.5690	5.5710
Modified *BIC*	5.6382	5.5803	5.5729	5.5761
CAIC	5.6383	5.5804	5.5730	5.5763

The table presents values of different information criteria for Markov switching models with 4 states and M clusters. For simplification, the information criteria are divided by the number of observations. Each information criterion prefers the model with the smallest value of the criterion.

the crisis state is not the negativity of the growth rate, but the instability thereof, which manifests itself both in the high standard error and in the negative autocorrelation of the growth rates. Despite these reservations we retain the crisis regime naming for ease of comparison. The second state corresponds to Jerzmanowski's stagnation state. The constant of this growth regime is not significantly different from zero and features significant but rather small positive autocorrelation. The standard error of this regime is much smaller than that of the instable growth regime. The implied steady state growth rate is close to zero. The third state is characterized by a steady state growth rate of 1.89% and a high persistence of growth rates (0.38). It corresponds to Jerzmanowski's stable growth regime, which for instance the industrialized countries spend most of their time in. The stability implied by this state is further corroborated by the small standard error associated with this regime (1.94%). This is the smallest standard error across all regimes. Finally, the fourth state captures rapid growth and corresponds to Jerzmanowski's miracle growth regime. It is characterized by a large positive constant, a relatively high autocorrelations and a relatively small standard deviation. The implied steady state growth rate equals 6.56%. This state is most often visited by the well known growth miracle countries such as Korea, Taiwan, Singapore or China. Other countries that have managed to catch up with developed countries such as the European success stories Greece, Portugal, Ireland or Spain, have also spent non-negligible amounts of time in this regime.

Table 5 shows the estimates for the transition matrices and the classification results for $M = 3$ clusters. The different transition matrices reflect the fact that even though the growth experience of all countries involves the same set of regimes, the combination of the states along the time-path varies distinctly. The estimated transition matrices imply quite diverse growth dynamics, as can most easily be seen by comparing the implied ergodic distributions.

In the long run, countries in the first cluster spend 80 % of their time in the stable growth state. Occasionally the stability of the economies is interrupted by a short recession or by a longer growth spurt. Serious drops in output or other forms of instable growth hardly ever occur. The prominence of the stable growth state is the consequence of its high persistence once it has been reached: countries remain in the stable growth regime with a probability exceeding 98 %. If countries expe-

Table 4: Parameter Estimates (Four Growth Regimes, Three Clusters)

	Constant	AR Coefficient	Standard Deviation	Steady State Growth Rate
State 1	−0.43	−0.5683*	13.89	−0.27
State 2	−0.04	0.0722*	4.70	−0.04
State 3	1.17*	0.3787*	1.94	1.89
State 4	4.92*	0.2494*	2.52	6.56

Except for the AR coefficients all numbers are percentages. The stars denote the significance of the estimate at the 5 % level.

rience a stagnation they remain in stagnation on average for two years. Chances are good that they enter fast growth afterwards and remain there for quite a long period of time (the probability of remaining in the rapid growth regime is 85 %) or that they return to the stable growth state. After a serious crisis, which according to the transition probability matrix hardly ever exceeds one year, the countries almost surely start to grow rapidly, thus making up for the income loss in a short time. In the lower part of table 5 the members of the first growth cluster are displayed subdivided by the geographic location. Essentially, the cluster consists of today's industrialized and newly developed countries. We find it surprising that a three cluster growth model does not subdivide these countries into different clusters even though the growth rates in the newly developed countries have been consistently much higher than those in the industrialized countries over the last forty years. The estimation algorithm perceives the distinction between the dynamics of the developing countries assembled in cluster two and three as more important. Today's industrialized and newly developed countries are separated only if the model is estimated under the assumption of four clusters. Apart from the industrialized and newly industrialized countries cluster one also contains some developing countries such as South Africa, Guatemala, El Salvador, Columbia or Paraguay. This feature follows from the fact that the growth regimes have been defined without consideration of countries' income levels. Such a definition entails that steadily growing developing countries will find themselves in the same cluster as steadily growing industrialized countries. While an extension of the states to include income levels is surely desirable, it would require longer time-series than those that are currently available. The main characteristic of the first cluster is its stability and the implied positive economic development. We will refer to this cluster as the successful growth cluster, because it includes both countries growing steadily and developing countries catching up with the world leader countries.

Compared to the first cluster, countries in the second cluster spend less time in the stable growth regime (45 %). Instead, they find themselves much more often in stagnation (37 %). In the long run, the countries grow rapidly for almost 17 % of their time and hardly ever find themselves in the crisis growth state. Focusing on the ergodic distribution of growth regimes only gives the impression that cluster one and cluster two countries are similar with regard to rapid growth. A

Table 5: Transition Probability Estimates and Country Classification

	Successful Growth Cluster					Moderately Successful Growth Cluster					Growth Failures			
	State 1	State 2	State 3	State 4		State 1	State 2	State 3	State 4		State 1	State 2	State 3	State 4
State 1	0.0000	0.0000	0.0023	0.0000	State 1	0.0000	0.0208	0.0206	0.0000	State 1	0.3331	0.1923	0.0000	0.1298
State 2	0.0010	0.4416	0.0094	0.1491	State 2	0.6146	0.6822	0.0713	0.4506	State 2	0.5199	0.7382	0.1900	0.3329
State 3	0.0000	0.1800	0.9862	0.0000	State 3	0.0000	0.0009	0.8101	0.5226	State 3	0.0000	0.0521	0.8099	0.1178
State 4	0.9990	0.3784	0.0020	0.8509	State 4	0.3854	0.2961	0.0980	0.0268	State 4	0.1470	0.0174	0.0001	0.4194
Ergodic Distribution	0.0018	0.0587	0.7682	0.1713	Ergodic Distribution	0.0169	0.3667	0.4525	0.1638	Ergodic Distribution	0.1759	0.5685	0.1939	0.0616

Africa	**North America**	**Africa**	**North America**	**Africa**	**North America**
South Africa	Canada	Burkina Faso	Costa Rica	Algeria	Nicaragua
	Guatemala	Egypt	Dominican Rep.	Benin	
Asia	El Salvador	Guinea	Honduras	Burundi	**South America**
Hong Kong	United States	Kenya	Jamaica	Cameroon	Argentina
Japan		Morocco	Mexico	Chad	Ecuador
South Korea	**South America**	Tunisia	Panama	Cote d'Ivoire	Peru
Malaysia	Brazil			Ethiopia	Venezuela
Singapore	Colombia	**Asia**	**South America**	Ghana	
Taiwan	Paraguay	India	Bolivia	Madagascar	
Thailand		Indonesia	Chile	Malawi	
	Oceania	Israel	Uruguay	Mali	
Europe	Australia	Nepal		Mozambique	
Austria	New Zealand	Pakistan		Niger	
Belgium		Philippines		Nigeria	
France		Sri Lanka		Rwanda	
Greece		Turkey		Senegal	
Ireland				Tanzania	
Italy		**Europe**		Togo	
Netherlands		Denmark		Uganda	
Norway		Finland		Zambia	
Portugal				Zimbabwe	
Spain					
Sweden				**Asia**	
Switzerland				Iran	
United Kingdom				Syria	

In each transition matrix the entry in row j, column i should be interpreted as $p_{ij} = P(s_t = j | s_{t-1} = i)$.

closer look at the transition probability matrix, however, reveals, that while countries in cluster one experience persistent phases of rapid growth, countries in cluster two mainly experience fast growth for a very short period of time (the probability of remaining in the miracle growth regime once it has been reached is less than 3 %). Put more provocatively, rapid growth in cluster one generates catching up with more developed countries along with the required extension of production capacities whereas rapid growth in cluster two is a one-off event which most often occurs after a period of stagnation or severe crisis. It essentially reflects the renewed utilization of the existing productive capacities, but not the expansion of them. The countries included in this cluster are unspectacular developing countries: they are neither known as success stories nor as prominent failures. Somewhat surprisingly, Denmark and Finland are assigned to this cluster as well.

The dynamics of the final cluster are the least favorable ones from a growth perspective. In the long run, the countries associated with this cluster spend 75 % of their time either in stagnation or in crisis. Steady growth accounts for 19 % of the time while rapid growth occurs only occasionally (6 %). An inspection of the transition probability matrix reveals that from every state there is a high probability of returning either to stagnation or of falling into crisis. Countries in this cluster on average experience an ever widening income gap to the developed countries because their average growth rate falls short of the 2 - 3 % growth rate of industrialized countries. The amount of time these countries spend in rapid growth does not make up for the long periods of time during which they stagnate or even shrink. The cluster contains the majority (75 %) of the African countries in the sample, which conforms to the perception of growth in Africa being a "tragedy" (Easterly and Levine, 1997). Other countries known for their erratic growth behaviour and erratic economic policies such as Argentina or Venezuela, also belong to this cluster.

The preceding clustering results corroborate Paap et al. (2005)'s finding that a priori clustering such as the use of geographic dummy variables tends to oversimplify the clustering structure. For instance, a dummy variable for Africa would disregard the fact that even though 75 % of the African countries indeed belong to the growth failure cluster, 25 % of the countries do not. It is problematic to group all African countries into one cluster, because this grouping also implies parameter homogeneity across the countries. However, in the Markov switching model policy or external shocks cause reactions that are governed by the underlying transition probability matrix. Hence, if countries belong to different clusters, their reaction towards shocks cannot be homogeneous. It follows that parameter homogeneity might be too strong an assumption to make and leads to potentially misspecified transmission mechanisms.

To end this section we want to shortly reflect on the main difference between our approach and that by Jerzmanowski (2006). Recall that Jerzmanowski estimates the transition probability matrix conditional on the institutional quality. Since all old industrialized nations possess institutions of

comparable quality, in the latter approach it could never be the case that two old industrialized nations like Finland and Denmark have very different transition matrices from the other nations. Since our approach is purely driven by the observed growth processes, such an unexpected assignment can happen. In essence, by conditioning the transition probability matrices on the institutional quality measured in a particular year, the state of the economy in that period is implicitly given more weight than the states of the economy in other years in Jerzmanowski's approach. Contrary to that, the approach used in this paper weights the growth experience of all years equally.

4.2   Growth Processes

In this section the implications of the parameter estimates for the growth processes of different countries are studied in more detail. As a first step, the average time an economy spends in the different regimes is uncovered. Following this, the importance of accounting for the different transition probability matrices is exemplified by some counterfactual cluster classifications. Finally, the correspondence of growth regime switches with actual political and economic events is established.

Table 19 reports the average smoothed regime probabilities for selected countries. The average smoothed regime probabilities indicate the probability that a country is in a certain regime in an average year. Not surprisingly, countries that have exhibited poor growth performance with a negative overall growth rate on average spend more time in the crisis state than more successful countries. Accordingly, miracle growth countries on average spend a lot of time in the growth miracle regime, e. g. China and Korea. However, table 19 also corroborates Pritchett's observation that in general the average growth rate masks quite distinct growth dynamics. If the dynamics of growth are of interest in themselves, for instance because erratic growth implies welfare losses due to the heightened uncertainty, then it is certainly not enough to focus on the average growth rate only (Becker and Mauro, 2006; Pritchett, 2000). Consider the group of countries growing with an average growth rate around 2.65 %. France and to a lesser extent Italy have done so by remaining in the stable growth state mainly, which is reflected in the high average regime probability of remaining in the stable growth state. Finland and Israel have a more mixed history: on average they have spent one third of their time in the stagnation state. The lower growth during these periods has been made up by achieving rapid growth in 17 % of their time. An even more extreme example is Ghana. Ghana has largely spent its time stagnating or being in crisis. It's relatively high average growth rate is achieved by going through extremes: its growth rates often jump erratically between double-digit negative and double digit-positive. The example of Ghana emphasizes that being in the crisis regime should not be mistaken as necessarily meaning negative growth: the crisis state can also capture unsustainably high growth rates. Similar stories apply throughout: Burkina Faso, Malawi and El Salvador all grew on average by around 1.25 %. But whereas growth in El Salvador has been smooth (92 % of the time is spent in the stable growth state), Burkina Faso achieved the growth rate despite frequent periods of stagnation thanks to relatively frequent periods of miracle

Table 6: Average Regime Probabilities

Country	Cluster	Crisis	Stagnation	Stable Growth	Miracle Growth	Average Growth Rate
Chad	3	0.25	0.70	0.02	0.03	-0.68
Senegal	3	0.15	0.72	0.09	0.05	-0.48
Togo	3	0.15	0.70	0.07	0.08	-0.40
Burkina Faso	2	0.01	0.65	0.15	0.19	1.24
El Salvador	1	0.00	0.07	0.92	0.00	1.18
Malawi	3	0.15	0.48	0.31	0.06	1.29
Finland	2	0.01	0.28	0.54	0.17	2.66
France	1	0.00	0.00	0.99	0.00	2.69
Ghana	3	0.36	0.53	0.09	0.03	2.74
Israel	2	0.03	0.39	0.41	0.17	2.62
Italy	1	0.00	0.08	0.78	0.13	2.75
Greece	1	0.00	0.08	0.61	0.30	3.11
Norway	1	0.00	0.00	1.00	0.00	3.11
Japan	1	0.00	0.05	0.70	0.26	3.97
Portugal	1	0.00	0.12	0.57	0.31	3.79
Sri Lanka	2	0.01	0.26	0.52	0.21	3.83
Malaysia	1	0.02	0.16	0.23	0.58	4.77
Singapore	1	0.00	0.23	0.02	0.75	4.72
China	1	0.00	0.16	0.00	0.84	6.69
South Korea	1	0.00	0.13	0.03	0.84	6.35

In this table, the average smoothed state probabilities, i. e. the probability for a country to be in a certain regime in an average year are reported. Formally, each column equals $(1/T)\sum_{t=1}^{T} P(s_t = j | \psi_T)$, where $P(\cdot)$ is the estimated probability of country k being in state j given the information of the entire sample and conditional on its cluster assignment.

growth. Malawi, on the other hand, achieved the growth rate by growing steadily for one third of the time and by visiting the crisis state and profiting from some extremely high growth rates. Norway and Greece are characterized by an average growth rate of 3.11 %, but whereas Norway grew smoothly, Greece went though a catch-up process. Singapore achieved its yearly average growth rate of 4.72 % by being in the miracle growth state for three-quarters of the time. The same average growth rate has prevailed in Malaysia, but has been generated by less stagnation, more stable and less miracle growth. It should be noted that the differences in the average smoothed regime probabilities do not only arise because countries find themselves in different clusters. A large number of the differences are observable despite countries being in the same cluster (e. g.

Greece and Norway, Singapore and Malaysia).[28]

Table 7: Counterfactual Smoothed State Probabilities

Country	Cluster	Crisis	Stagnation	Stable Growth	Miracle Growth
China	**1**	0.00	0.16	0.00	0.84
	2	0.02	0.43	0.24	0.30
	3	0.12	0.30	0.03	0.55
France	**1**	0.00	0.00	0.99	0.00
	2	0.00	0.04	0.89	0.07
	3	0.00	0.05	0.94	0.01
Ghana	1	0.08	0.66	0.15	0.12
	2	0.19	0.59	0.11	0.11
	3	0.36	0.53	0.09	0.03
Malawi	1	0.01	0.37	0.47	0.16
	2	0.05	0.48	0.34	0.13
	3	0.15	0.48	0.31	0.06
Malaysia	**1**	0.02	0.16	0.23	0.58
	2	0.03	0.23	0.56	0.19
	3	0.07	0.33	0.42	0.18
Norway	**1**	0.00	0.00	1.00	0.00
	2	0.00	0.04	0.89	0.08
	3	0.00	0.04	0.95	0.01
Sri Lanka	1	0.00	0.24	0.26	0.50
	2	0.01	0.26	0.52	0.21
	3	0.07	0.42	0.42	0.09

This table reports both the actual and the counterfactual average smoothed regime probabilities for selected countries. The actual regime probabilities are indicated by the bold cluster digit.

So how important is the assignment to clusters? One way to assess the importance is to look at how the average smoothed regime probabilities change from the real to counterfactual cluster assignments, which is shown for a selection of countries in table 7. It transpires that the impact of the assignment depends on the stability of the growth rate series. For instance, France and Norway are characterized by very stable growth processes, remaining almost all the time in the stable growth regime when classified into cluster one. Even if these countries are classified into different clusters, the essence of stability continues to be captured because in all clusters the stable growth

---

[28] A table containing the average regime probabilities for all countries can be found in appendix C.

regime is quite persistent once it has been reached. However, if countries are characterized by more volatile or extreme growth processes, the cluster assignment becomes more important. Consider for instance China and Malaysia, two miracle growth countries. If these countries are assigned to cluster one, most of their growth behavior is interpreted as arising from the miracle growth state. If, however, they are assigned to cluster two or cluster three, the time spent in the miracle growth regime is diminished considerably, simply because the transition matrices of cluster two and three feature much less persistence of the miracle growth state than the transition matrix of cluster one. In a similar vain, if countries exhibit highly erratic growth behavior, they are usually assigned to cluster three, so that a relatively large fraction of their time is spent in the crisis state (e. g. Ghana or Malawi). If these countries are counterfactually assigned to clusters one or two, much more of their growth is assigned to stagnation or miracle growth simply because these transition matrices do not allow for persistent crises or stagnations. Hence, the assignment to the clusters has a bearing on the interpretation of the growth process. It becomes more important the more distinctive the growth processes under consideration are.
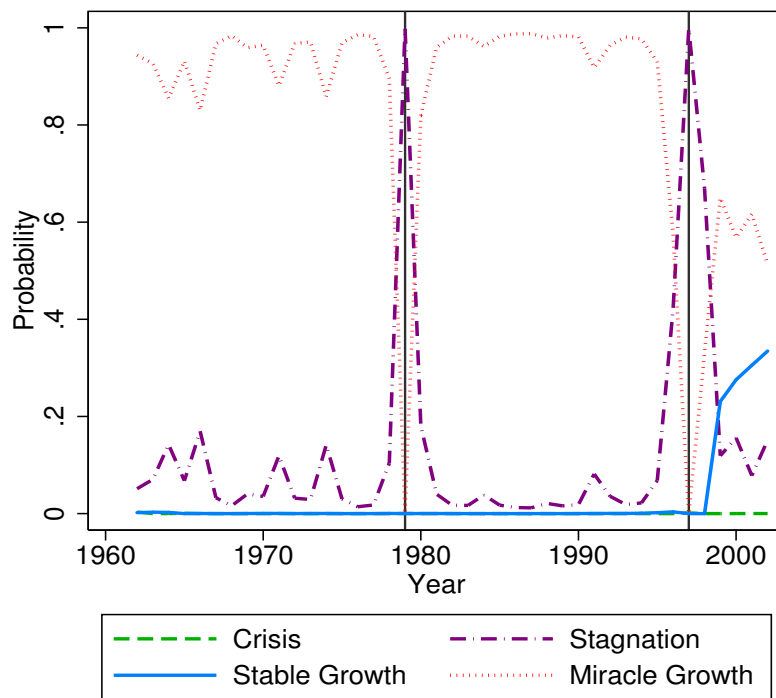
Figure 1: Average Regime Probabilities: Korea

As a final step, the question whether identified growth regime changes correspond to important economic or political events in the respective economies is examined. Figure 1 plots the smoothed state probabilities for Korea, a well known growth miracle. Not surprisingly, in almost all periods the miracle growth state clearly dominates all other states. However, in 1979 and 1997 the probability of stagnation dominates. These regime changes can easily be related to political and economic events. In 1979 president Park Chung-Hee was assassinated, which lead to a year of

political turmoil. 1997 marks the Asian Financial Crisis that had severe short-run impacts on the economy of South Korea. The next figure plots the smoothed regime probabilities for Japan. Unlike Korea, Japan underwent a permanent regime change. Before 1973 Japan has found itself in the miracle growth state. In 1973 Japan was hit hard by the first oil crisis, which shows in the stagnation regime becoming dominant. Afterwards, the growth rates have not reached the high pre-crisis levels again. Instead, Japan started to grow steadily within the stable growth regime. Unfortunately, our model is not sensitive enough to pick up the prolonged recession following the banking crisis in the 1990s. This failure points at one important shortcoming of the present model: the countries are assigned to clusters once and for all whereas a more realistic model would and should allow the transition probability matrices to change over time. After all, this is what today's developing countries are aiming at. They aim to achieve similar growth rates as the miracle growth countries did in the past. This in any case requires them to undergo reform processes that improve their transition probability matrices. Future research should therefore concentrate on making the transition probability matrices time-varying, something which neither this paper nor Jerzmanowski (2006) have done.

The correspondence between growth regime changes and actual events is not confined to countries in cluster one. Consider for instance Mexico (figure 3), a country belonging to cluster two. Mexico has been growing steadily until 1981, where a prolonged phase of stagnation lasting until 1987 set in. In the beginning of the eighties Mexico among other Latin American countries was
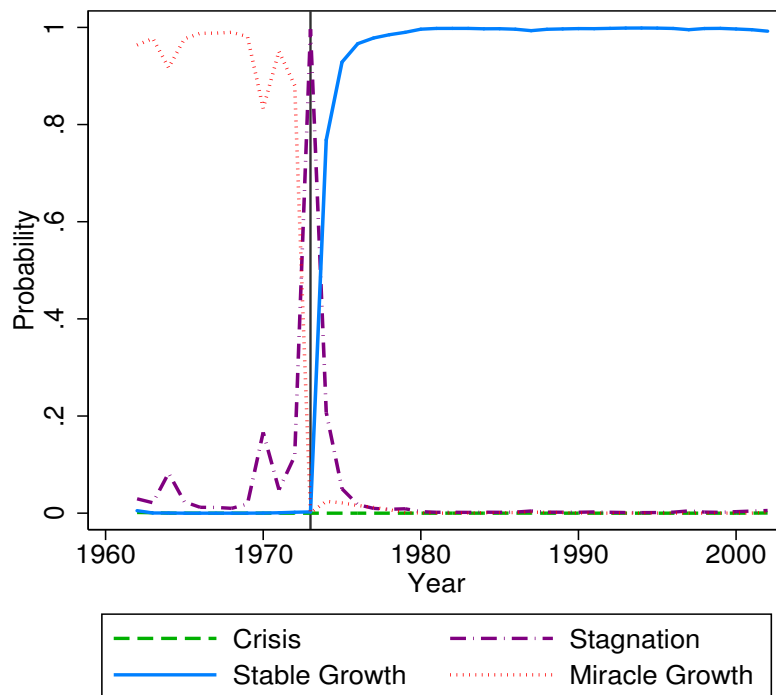

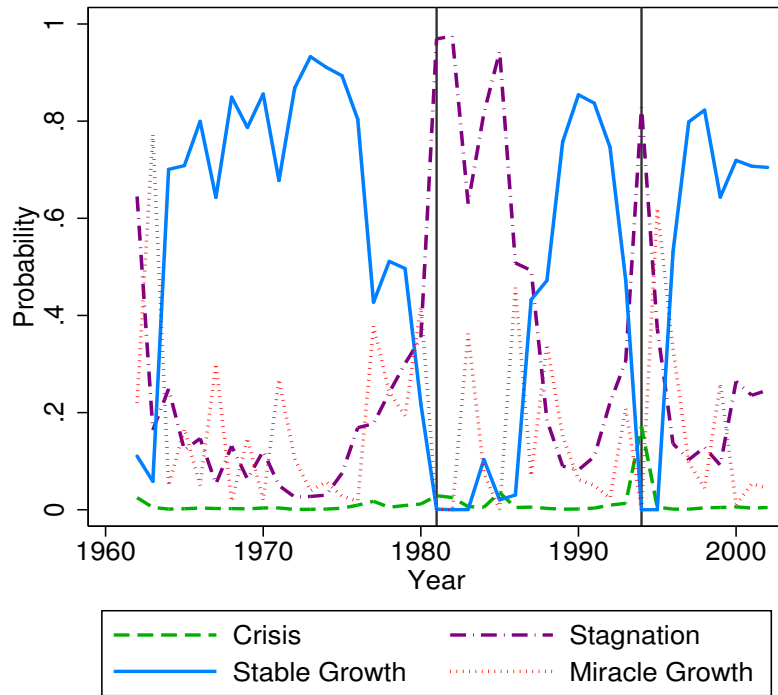Figure 2: Average Regime Probabilities: Japan

28

Figure 3: Average Regime Probabilities: Mexico

hit by the Latin American debt crisis, which led to the prolonged recession that is revealed in the smoothed state probabilities from 1981 to 1987. After some years of reasonable economic growth the smoothed regime probabilities also pick up the Mexican currency crisis in 1994. They indicate a recession lasting approximately two years. After that growth resumed in the stable growth regime again.

Algeria is provided as an example of countries in cluster three. Algeria spent most of its time in stagnation interrupted by periods of crisis in 1965, 1971 and 1978. The country entered a stable growth period only in 1994. All these regime changes can be linked to actual political or economic events. In 1965 Boumediene took over as president in a military coup d'état and remained in power until his death in 1978, which caused serious struggles about the successor. 1971 was characterized by the confiscation of French energy operations by the state. In the seventies and the eighties the economic situation in Algeria worsened and in the beginning of the 1990 the country had more and more difficulties meeting its external debt payments. In 1994 Algeria arranged a comprehensive debt rescheduling with the support of the IMF and introduced several important reforms as a results of this package, among them for instance trade liberalization . The measures were successful and helped Algeria to embark on a sustainable growth path at least until the end of our sample period.
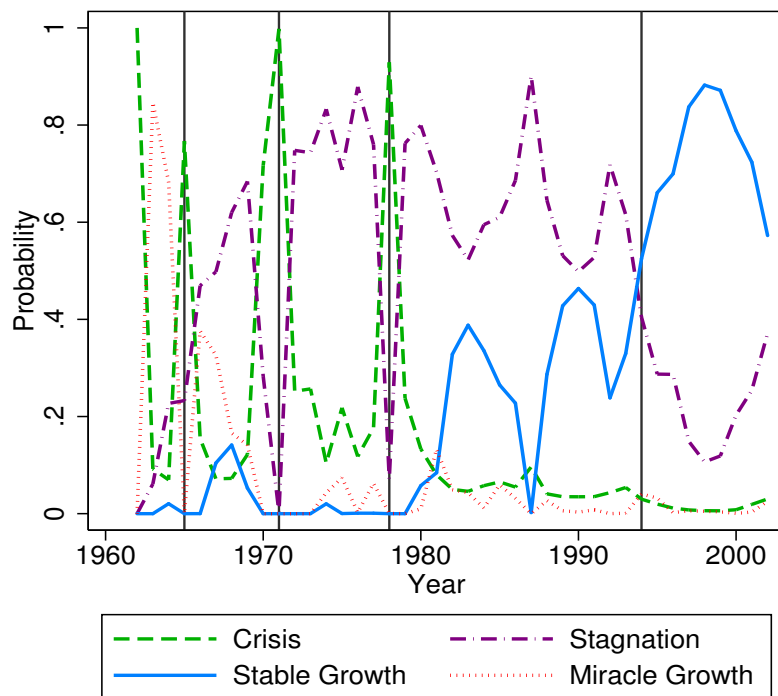
29

Figure 4: Average Regime Probabilities: Algeria

5 Accounting for the Assignment to Clusters

5.1 Conceptual Issues

In the previous section we have identified groups of countries that follow the same growth process in the sense that the growth path is generated by the same Markov switching model. Each cluster implies a distinctive long-run growth performance: Countries in the first cluster experience more successful growth histories than countries in the second cluster, and these countries again are much more successful in achieving a "desirable" growth path than countries in the third cluster. It would obviously be interesting to understand the underlying forces that lead to such different transition matrices and growth dynamics. As Jerzmanowski (2006) argues, the driving forces behind such fundamental differences have to be "growth fundamentals" as opposed to short-term shocks or policy events. The Markov switching classification approach has the virtue that it allows comparisons across clusters and can thus be extremely helpful in identifying the relevant growth fundamentals. Jerzmanowski (2006) had to postulate an underlying growth dynamics - growth fundamentals relationship depending on one variable only because of the high non-linearity of the estimation problem. The use of the clustering technique allows us to investigate this relationship in a second step. Since the non-linearity of the estimation is no longer an issue in the second step, the relevance of multiple growth fundamentals can be analyzed. We follow the literature in applying a two-step procedure without making adjustments to account for the uncertainty inherent in the clustering.[29] Therefore, our analysis can only be seen as a first step to uncover the driving forces for different growth dynamics. Nevertheless, it might prove a valuable starting point for further research.

In an interesting contribution on the relationship between multiple growth accelerations and the success of countries to jump out of underdevelopment traps, Berthelemy (2006) has provided a comprehensive list of growth fundamentals. He divides the factors that may help in escaping an underdevelopment trap into three broad categories: The first category relates to factor accumulation, the second one to the structure of the domestic economy and the third one to the prevailing political institutions. The first category is based on the idea that there are non-convexities in the production process and therefore thresholds below which an economy is caught in an underdevelopment trap. One potential threshold relates to the capital stock per worker. Capital accumulation may be hampered to exceed certain levels because population growth inhibits the per capita accumulation of capital (Nelson, 1956) or because meeting subsistence consumption prevents saving and investment (**?**). Besides a minimal level of physical capital per capita to jump start growth, Azariadis and Drazen (1990) point out that a minimal level of human capital might be equally important for increasing returns to scale in production to materialize. Countries may also escape underdevelopment traps because they succeed in creating favorable structural features in their economies.

---

[29] For instance, Durlauf and Johnson (1995) and Ardic (2006) have followed this approach.

For instance, they might implement a rather strong financial system, thus facilitating the capital accumulation process (King and Levine, 1993b,a). The structural features might also be related to geographic conditions such as climate, access to the sea or rainfall (Bloom et al., 2003). Majumdar and Mitra (1995), and Dasgupta (1998) show that countries may escape underdevelopment traps by integrating into the world economy. In addition, appropriate political institutions can facilitate or hamper the escape from an underdevelopment trap. Collier (2006) argues in detail that the lack of appropriate institutions in Africa is responsible for its dismal growth record. Acemoglu et al. (2001, 2003) argue that economic growth, the level of development and macroeconomic volatility are all ultimately determined by the institutional environment of the economy.

In the following, the explanatory power of each category for the assignment of countries to clusters will be analyzed. In the first part of the analysis we will look for significant differences in the respective categories at the beginning of the sample period across clusters without accounting for interdependencies. Using these results a parsimonious logit model will be fitted. The focus is on the beginning of the sample period because otherwise it becomes impossible to distinguish whether the observed conditions are causing the assignment to clusters or whether the growth experiences resulting from the assignment to a cluster are causing the observed conditions to change.[30] Even the validity of assuming the initial conditions to be exogenous or predetermined can be questioned: The clustering procedure assumes that countries belong to one cluster once and for all, which implies that the countries have been members of the respective clusters also before 1960. It follows that even the initial conditions might be subject to an endogeneity bias leaving only physically given constraints as valid targets of the investigation. However, it is common practise in growth econometrics to treat initial conditions as exogenous. Therefore, we, too, are going to abstract from this issue.

Throughout the analysis those countries that had already been well developed in 1960 will be separated into a different cluster from the other countries that still had to develop in 1960. The developed countries are defined as those that today belong to the OECD countries and whose income per capita in 1960 has been at least half the income per capita in the United States. The developed countries are assigned to the new cluster 4, which consists of Australia, Austria, Belgium, Canada, Denmark, Finland, France, Italy, Norway, New Zealand, The Netherlands, Sweden, Switzerland, the United Kingdom and the United States of America. In the following these countries are called the industrialized or developed countries. There are two reasons for this procedure. The first reason

---

[30] The small number of countries prevents the use of instrumental variable or control function strategies as proposed by Wooldridge (2002, chap.15.7.2), and Imbens and Wooldridge (2007), because these techniques are only valid asymptotically. Moreover, in the present context it is questionable whether there are any valid instruments that meet the requirement to be correlated with the endogenous variable and at the same time to be uncorrelated with any other omitted variable that is potentially useful in explaining the assignment to clusters (Brock and Durlauf, 2001).

is pragmatic: the developed countries differ significantly in almost all respects from the developing countries so that averages across these subgroups are not meaningful for either of the countries. Moreover, from a development perspective it is more interesting to learn what developing countries can do right instead of confirming that industrialized countries have indeed been different from developing countries. It is by now widely accepted that it is too much to ask for developing countries to implement first-best solutions as a growth strategy. Rather, it is argued that the focus should shift towards identifying the binding constraints for growth and alleviating these constraints (Hausmann et al., 2004). Therefore, the differences between successful and less successful developing countries at the beginning of the sample period have more potential to generate reasonable policy suggestions than the differences between industrialized and developing countries.

5.2   A Descriptive Approach

As a first step, the investment and saving rates across clusters are analyzed in order to assess whether countries in less successful growth clusters have suffered from an insufficient amount of physical capital due to low saving or investment propensities. Both the saving and the investment rates should convey the same message because except for external finance decisions savings should equal investments. However, for developing countries foreign direct investments or foreign aid might contribute significantly to the country's investment possibilities and capital drains might significantly decrease them. The saving and investment rates are reported for the years 1960, 1965 and 1970 because for many countries the respective series only start at a later point in time. For each pair of clusters group mean comparison tests are conducted assuming unequal variances using Satterthwaite's degree of freedom adjustment. If the difference between two clusters is significant on the five percent level, this is indicated by a superscript. For reasons of clarity the significance of the differences for each cluster are only tested against clusters with a larger number. However, obviously if the difference of means between cluster 1 and cluster 4 is significant, so is the difference of means between cluster 4 and cluster 1. A significant difference should only be interpreted as loosely indicating the potential relevance of the variable and not as formal evidence. Even if a significant difference at the five percent level were found for each pair of clusters, it does not automatically follow that the variable is significant overall, because the repeated testing procedure inflates the type I error rates.[31]

Table 8 reveals that the differences between the saving and the investment rates are not very pronounced between the development countries aka 1960. The investment rates of the newly developing countries in cluster one only becomes significantly different from those in less developing

---

[31] For instance, according to the simple Bonferroni adjustment for the overall type I error rate to be less or equal to five percent, each individual variable would have to be significant at the 0.05 / 6 = 0.0083 significance level (Shaffer, 1995).

Table 8: Saving and Investment Rates

	Investment Rate			Saving Rate		
	1960	1965	1970	1960	1965	1970
Cluster 1	17.15[4]	19.42[4]	21.57[2,3]	19.72	20.92	23.59[2,3]
	(2.40)	(2.54)	(2.76)	(3.27)	(2.78)	(2.57)
	18	18	18	13	14	15
Cluster 2	11.97[4]	13.31[4]	14.36[4]	14.31	14.23[4]	13.76[4]
	(1.57)	(1.67)	(1.41)	(1.56)	(1.83)	(1.68)
	22	23	23	15	18	20
Cluster 3	12.71[4]	12.93[4]	12.62[4]	14.92	14.63[4]	16.73[4]
	(3.14)	(3.21)	(3.40)	(2.86)	(2.47)	(2.32)
	28	28	28	20	23	24
Cluster 4	24.90	25.55	26.08	25.58	26.51	27.39
	(1.38)	(1.15)	(1.45)	(4.19)	(1.77)	(1.57)
	15	15	15	4	8	12

For each cluster the mean, standard error of the mean and the number of observations is reported. The superscript i attached at entry j indicates a significant difference between the means of cluster i and cluster j at the five percent level using Sattersthwaite's degree of freedom adjustment. The investment rates is taken from the Penn World Tables 6.2 (Heston et al., 2006), the saving rate from the World Development Indicators 2007 (WorldBank, 2007).

successful countries in 1970. Hence, the improvement may easily be a consequence of the growth experiences up to that point. The saving rates between the clusters are even less distinctive. The industrialized countries, however, are characterized by significantly higher investment rates (around 25 %) and significantly higher saving rates (also around 25 %).[32] It thus appears that the initial difference between saving and investment rates are not well suited to explain the different growth dynamics of developing countries.

Even if the saving and investment rates do not differ significantly across clusters, it might be that similar investment rates influence economic development differently because capital is diluted across persons due to high population growth rates. It is also conceivable that the physical capital stock in some countries is subject to decreasing returns to scale because there are not enough young and healthy people to operate on it. For these reasons demographic statistics are reported in table 9. A similar picture as with the investment and saving rates emerges. In 1960 only the developed subcluster stands out, having a much lower dependency ratio and lower population growth rates than the other clusters. Until 1970 the differences between the growth failure and the other clusters widen in that cluster three has a noticeably higher population growth rate and dependency ratio than all the other clusters. Nevertheless, in sum the demographic development does not show much promise in explaining the different growth dynamics.

---

[32] We attribute the insignificant difference in the saving rates of 1960 to the small number of observations.

Table 9: Demographic statistics

	Population Growth Rate			Dependency Ratio		
	1960	1965	1970	1960	1965	1970
Cluster 1	2.47^4	2.15^4	$1.94^{3,4}$	$0.80^{3,4}$	$0.81^{3,4}$	$0.79^{2,3,4}$
	(0.36)	(0.25)	(0.24)	(0.04)	(0.04)	(0.04)
	18	18	18	17	17	17
Cluster 2	2.48^4	2.49^4	$2.37^{3,4}$	0.85^4	0.89^4	0.90^4
	(0.15)	(0.13)	(0.14)	(0.02)	(0.03)	(0.03)
	23	23	23	23	23	23
Cluster 3	2.53^4	2.66^4	2.83^4	0.89^4	0.92^4	0.93^4
	(0.12)	(0.11)	(0.12)	(0.02)	(0.02)	(0.02)
	28	28	28	28	28	28
Cluster 4	1.17	1.06	0.87	0.59	0.59	0.59
	(0.17)	(0.12)	(0.17)	(0.02)	(0.02)	(0.01)
	15	15	15	15	15	15

For each cluster the mean, standard error of the mean and the number of observations is reported. The superscript i attached at entry j indicates a significant difference between the means of cluster i and cluster j at the five percent level using Sattersthwaite's degree of freedom adjustment. The population growth rate is taken from the Penn World Tables 6.2 Heston et al. (2006), the dependency ratio is calculated using the World Development Indicators 2007 (WorldBank, 2007).

Unlike the demographic variables educational achievements vary distinctly across clusters. Two indicators for the average years of schooling of the population over the age of 15 are reported: the first is taken from Cohen and Soto (2001) for the year 1960 and the second is taken from Lutz et al. (2007) for the year 1970, both of which represent the starting years of the respective data series. The second series is derived using demographic multi-state methods to project backwards information from an empirical dataset collected in 2002 whereas the first series attempts to use as much observable data and as little extrapolation as possible in order to come up with the estimates for the average years of schooling per person at the age of 15 years or above.[33] Both data sets tell the same story: Industrialized countries had by far the best educated work force with an average education of 8 years. The newly developed countries follow with almost 4 years of education, the less successful countries had around 3 years of education whereas in the least successful countries the average person possessed less than 2 years of education. According to the Cohen-Soto data the means of all clusters differ significantly from each other, whereas according to the Lutz data only the industrialized and the least successful countries stand out. The average share of the population having completed at least primary education in 1960 is reported as well (Barro, 2000; Barro and Lee, 2001). This measure, too, indicates that the group means in 1960 are significantly different from each other for all cluster pairs. These observations provide some support to the idea that sluggish development can be related to too little education and that this variable has the potential

---

[33] If a country has only been available in one of the two datasets, the data has been extrapolated using linear regression in order to achieve as much data coverage as possible.

Table 10: Educational Statistics

	Years of Schooling 1960	Years of Schooling 1970	Primary School Completed 1960	Change in Years of Schooling 1960-2000	Change in Years of Schooling 1970-2000	Change in Primary School Completed 1960-2000
Cluster 1	$3.77^{2,3,4}$	$3.89^{3,4}$	$39.36^{2,3,4}$	3.92^{3}	$3.86^{2,3,4}$	23.76^{4}
	(0.43)	(0.36)	(4.11)	(0.35)	(0.25)	(2.82)
	18	18	18	18	18	18
Cluster 2	$2.50^{3,4}$	$3.31^{3,4}$	$25.88^{3,4}$	3.16^{3}	2.89	23.48^{4}
	(0.41)	(0.48)	(3.69)	(0.26)	(0.19)	(3.09)
	23	23	21	23	23	21
Cluster 3	1.43^{4}	1.89^{4}	13.73^{4}	2.40^{4}	2.66	17.64
	(0.24)	(0.29)	(1.68)	(0.24)	(0.19)	(2.76)
	28	28	25	28	28	25
Cluster 4	7.81	8.87	75.57	3.64	2.22	9.40
	(0.34)	(0.66)	(2.86)	(0.15)	(0.34)	(3.14)
	15	15	15	15	15	15

For each cluster the mean, standard error of the mean and the number of observations is reported. The superscript i attached at entry j indicates a significant difference between the means of cluster i and cluster j at the five percent level using Sattersthwaite's degree of freedom adjustment. Years of Schooling in 1960 are taken from ?, years of schooling in 1970 from Lutz et al. (2007) and the percentage of population having completed at least primary school from ?Barro and Lee (2001).

to explain the assignment of countries to clusters. Out of interest, we also report the average change in the educational measures from 1960/1970 to 2000. The t-tests for these measures suggest that the least successful countries have achieved a significantly lower expansion of the human capital stock than the other clusters. According to the Lutz data the newly developed countries have expanded their human capital stock significantly more than all the other clusters. With regard to the proportion of the population having completed primary education, all clusters have improved their situation. Obviously, due to the already large proportion of educated people the improvement is the smallest for industrialized countries.

Next, some structural features across clusters are compared. We start by looking at the financial structure of the economies. The availability of financial services can be measured by the ratio of liquid liabilities (M3) to GDP or by the ratio of money and quasi money (M2) to GDP. The conclusions are similar so for brevity only the first indicator is presented.[34] The implicit assumption behind these measures is that the size of the financial sector in a country is positively related to the provision of financial services (King and Levine, 1993b,a). A more direct measure of the provided financial services is the amount of financial resources directed to the private sector as a percentage of GDP. Table 11 indicates that the financial sector was significantly more developed in industrialized

---

[34] In contrast to "M2 to GDP" the indicator "M3 to GDP" also includes commercial papers and thus captures the development of the financial sector somewhat more comprehensively.

Table 11: Financial statistics

	M3 to GDP			Credit to GDP		
	1960	1965	1970	1960	1965	1970
Cluster 1	$21.34^{3,4}$	$31.91^{3,4}$	40.31^{3}	$21.70^{3,4}$	29.33^{3}	$38.67^{2,3}$
	(4.10)	(6.67)	(8.45)	(4.07)	(5.78)	(6.90)
	9	11	11	13	15	15
Cluster 2	$20.79^{3,4}$	$22.13^{3,4}$	25.19^{4}	14.36^{4}	16.68^{4}	$19.55^{3,4}$
	(2.55)	(2.37)	(2.32)	(2.07)	(2.07)	(2.20)
	17	21	22	17	20	21
Cluster 3	10.56^{4}	15.81^{4}	20.38^{4}	11.88^{4}	11.93^{4}	13.85^{4}
	(2.33)	(1.57)	(2.28)	(2.07)	(1.34)	(1.52)
	16	22	24	13	23	24
Cluster 4	52.65	56.00	59.35	39.31	44.30	49.92
	(5.20)	(4.54)	(5.44)	(6.61)	(6.54)	(6.72)
	11	13	13	13	15	15

For each cluster the mean, standard error of the mean and the number of observations is reported. The superscript i attached at entry j indicates a significant difference between the means of cluster i and cluster j at the five percent level using Sattersthwaite's degree of freedom adjustment. Both variables are taken from the World Development Indicators 2007 (WorldBank, 2007). The entry M3 to GDP in 1965 for cluster 2 is calculated excluding Ghana, because M3 to GDP was extraordinarily large (1997.9269), which seems to be a recording mistake.

countries at the beginning of the sample period than in other countries. Moreover, there is some evidence of the financial markets being much less developed in the least successful countries. The differences in the financial structure are not very pronounced between successful and moderately successful developing countries. One striking feature regarding the financial indicators is the high standard error of the mean in cluster one, indicating a potential outlier problem. The high standard error is caused by Japan, which displays much more developed financial markets than the other countries in cluster one. However, if Japan is dropped from the sample, it is still the case that only the underdevelopment of the third cluster is striking whereas the differences between clusters one and two remain insignificant. The observations might indicate that a minimum level of financial development is necessary to enable growth. However, a good financial structure is not what is driving the different growth dynamics between cluster one and two.[35]

The next structural feature to be analyzed refers to the integration of the countries into the world economy. It is commonly believed that the newly developed countries have benefited from openness and export-led growth. One particular problem with this feature is that the available data in general only indicates whether a country is open or not, and not the degrees of openness at the beginning of the sample period. Moreover, a large number of countries have liberalized in the Kennedy (1964-1967) and the Tokyo (1973 - 1979) GATT rounds, so that the value of an

---

[35] Berthelemy (2006) draws a similar conclusion when examining multiple versus single growth accelerations. He concludes that financial development is a consequence of rather than a cause for growth.

	Years of Liberalized Trade 1960 - 2002	Openness Dummy 1960	Openness Dummy 1965	Openness Dummy 1970
Cluster 1	$28.61^{2,3,4}$	0.39^{4}	$0.50^{2,3,4}$	$0.61^{2,3,4}$
	(3.40)	(0.12)	(0.12)	(0.12)
	18	18	18	18
Cluster 2	$15.04^{3,4}$	0.17^{4}	0.09^{4}	0.13^{4}
	(1.90)	(0.08)	(0.06)	(0.07)
	23	23	23	23
Cluster 3	7.61^{4}	0.14^{4}	0.11^{4}	0.04^{4}
	(1.46)	(0.07)	(0.06)	(0.04)
	28	28	28	28
Cluster 4	40.00	0.87	0.93	0.93
	(1.73)	(0.09)	(0.07)	(0.07)
	15	15	15	15

For each cluster the mean, standard error of the mean and the number of observations is reported. The superscript i attached at entry j indicates a significant difference between the means of cluster i and cluster j at the five percent level using Sattersthwaite's degree of freedom adjustment. The data is taken from Wacziarg and Welch (2008).

openness indicator is heavily dependent on the period it is measured. In addition to that, the causality between liberalization and growth is unclear. For these reasons we report the average years a country has been liberalized since 1960 to 2002 based on Wacziarg and Welch (2008) and the state of the economies in 1960, 1965 and 1970 in table (12). If a country has been open in a given year, the openness dummy has recorded a one. A clear trend is visible. The industrialized countries have been open for the whole time period, the newly industrialized countries for almost three quarters of the period. Contrary to that, the countries failing to grow in a satisfactory way have remained closed for most of the time. The greater openness of the industrialized and newly developed countries is also reflected in the dummy openness variables. Even if the causality issue is not entirely clear, openness of countries seems to be very useful in explaining cluster memberships. For the 1970 openness indictor it might be argued that it is reasonably exogenous because it lies close to the sample starting period and it therefore is unlikely that the growth experiences have already driven the policy decisions.

Geographic conditions also influence the structural features of an economy. For instance, access to the sea might make integration into the world economy easier and more worthwhile because transportation costs are lower. The climate or the availability of natural resources have a bearing on the production mix of an economy and the disease burdens (Gallup et al., 1999). In the following, the differences of the clusters with regard to access to the sea, the percentage of land and population within 100 km of the coastline and the percentage of land that is tropical are analyzed (see table 13). The most striking feature is that the least successful country cluster also has the

Table 13: Geographic Structure

	Landlocked Dummy	Percentage of Land Within 100 km to the Coast	Percentage of Population Within 100 km to the Coast	Percentage of Land in Tropical Climate
Cluster 1	0.06^3	0.59^3	0.70^3	$0.50^{3,4}$
	(0.06)	(0.09)	(0.08)	(0.11)
	18	18	18	18
Cluster 2	0.13	0.51^3	0.58^3	$0.58^{3,4}$
	(0.07)	(0.08)	(0.08)	(0.10)
	23	23	23	23
Cluster 3	0.36	0.14^4	0.27^4	0.85^4
	(0.09)	(0.03)	(0.05)	(0.06)
	28	28	28	28
Cluster 4	0.13	0.45	0.60	0.03
	(0.09)	(0.09)	(0.09)	(0.03)
	15	15	15	28

For each cluster the mean, standard error of the mean and the number of observations is reported. The superscript i attached at entry j indicates a significant difference between the means of cluster i and cluster j at the five percent level using Sattersthwaite's degree of freedom adjustment. All variables are taken from Gallup et al. (1999).

least favorable geographic conditions. 35 % of the countries do not have access to sea. Less than 30 % of the population lives close to the sea. However, 85 % of the land is tropical. The cluster means of these variables are all significantly different to those of the other clusters on the ten percent level suggesting some detrimental influence. On the other hand, there are landlocked or tropical countries in other clusters, too, so that it seems to be possible to overcome the disadvantages associated with these features. In sum, however, some geographic features seem to make growth harder to achieve than others so that the explanatory content of these variables for cluster assignment cannot be dismissed.

The final aspect to be analyzed is the influence of institutions on growth. Unfortunately, institutional quality has only been measured since the 1990 onwards on a large scale, so that allmost all measures refer to the end of the sample period raising the question of causality. The only exception is a variable compiled by the Business Environmental Risk Intelligence (BERI), which has been published in 1972 for the first time (Knack and Keefer, 1995).[36] Unfortunately, this index is only available for a limited number of developing countries in 1972, suggesting a possible sample selection bias. Apart from the BERI index, the government antidiversion index GADP created by Hall and Jones (1999), the voice and accountability indicator VA prepared by Kaufmann et al. (2008) and the judical indepences measure by ? are reported.

---

[36] We would like to express our thanks to Knack and Keefer (1995), who kindly provided their data collection.

Table 14: Quality of Institutions

	Government Antidiversion Index (GADP)	Voice and Accountability (VA1996)	Independence of Judicity	BERI index 1972	BERI index 1995
Cluster 1	0.69[2,3,4]	0.28[3,4]	0.69	10.51[2,3,4]	10.07[2,3]
	(0.04)	(0.17)	(0.08)	(0.54)	(0.60)
	17	18	13	12	13
Cluster 2	0.53[4]	-0.09[3,4]	0.79	6.74[4]	7.71[4]
	(0.02)	(0.14)	(0.08)	(0.74)	(0.43)
	22	23	13	11	8
Cluster 3	0.49[4]	-0.54[4]	0.73	7.22[4]	6.84[4]
	(0.01)	(0.14)	(0.09)	(0.55)	(0.58)
	28	28	15	5	5
Cluster 4	0.96	1.32	0.84	12.56	11.13
	(0.01)	(0.05)	(0.06)	(0.37)	(0.29)
	15	15	15	10	12

For each cluster the mean, standard error of the mean and the number of observations is reported. The superscript i attached at entry j indicates a significant difference between the means of cluster i and cluster j at the five percent level using Sattersthwaite's degree of freedom adjustment. The data is taken from Hall and Jones (1999), Kaufmann et al. (2008), ? and Knack and Keefer (1995).

The GADP index is based on the idea that a good social infrastructure is characterized by a tendency to keep the marginal and social returns of activities closely together. The index is an equal-weighted average of factors deemed important for this equality, namely of indicators for law and order, bureaucratic quality, corruption, risk of expropriation and government repudiation all taken from the International Country Risk Guide. The earliest available GADP indicator uses observations over the years 1986-1995. It is recorded as a number between zero and one, with a higher value denoting better institutions. The first column of table 14 reveals that the institutional quality has been significantly better in industrialized and cluster one countries, whereby the former significantly outperform the latter. The VA indicator summarizes the extent to which a country's citizens are able to participate in selecting their government and to which extent freedom of expression, association and media are granted. The value of the indicator can range between -2.5 and 2.5, whereby freedom is more restricted the lower the reported indicator. The results are similar to GADP: industrialized countries report the best VA indicator, the least successful countries the worst. Unlike the GADP indicator, the means of successful and moderately successful developing countries do not differ significantly. Judicial independence is the average of proxies for the tenure of Supreme Court judges, the tenure of administrative Court judges and for the prevalence of case law. It takes on the values 0, 1/3, 2/3 and 1, where 1 indicates judical independence. Unlike the previous two institutional variables, no significant differences across clusters can be observed. Finally, the results for the BERI indicator both for 1972 and 1995 are reported. This indicator aggregates aspects of contract enforceability, infrastructure quality, risk of expropriation without compensation and the efficiency of civil service and takes on values between 0 and 16. A higher value implies

better institutions. The results are similar to the GADP indicator: industrialized and cluster one countries have significantly better institutions than cluster two and cluster three countries. In 1972 the difference between cluster one and cluster four is significant whereas in 1995 it no longer is. While one is inclined to assume that this change is due to an improvement of institutions in cluster one, it is actually driven by a diminishing quality of the institutions in cluster four. Summing up, successful growth performance is associated with good institutions. However, institutional quality does not seem to be very helpful when the aim is to distinguish between countries with moderate and countries with no success in achieving growth.

5.3   A Multinomial Logit Approach

The multinomial logit model lends itself to summarize the foregoing discussion.[37] Selecting the appropriate model is particularly challenging in this setting because the number of observations overall and in every cluster is very small.[38] Hence, it is impossible to use the general-to-specific approach of model specification to arrive at the true model. Rather, a parsimonious model has to be specified even if this increases the danger of omitting relevant variables. Some recent results on logit specifications fortunately show that even though theoretically omitting a variable leads to inconsistent regressors, in practice prediction remains accurate and Wald tests of significance remain valid even though the power properties suffer somewhat. While the coefficient estimates are indeed subject to an attenuation bias and should not be directly interpreted, the magnitude of the effects can be interpreted by considering the average partial effects over the sample. Here the attenuation bias is mitigated by an inverse change of the probability density function (Cramer, 2007; Ramalho and Ramalho, 2007).[39]

According to the previous section education, openness and/ or geography and institutional structure have the most potential to explain the assignment to clusters. Therefore, our parsimonious base specification of the model contains these three variables. Suppose that the independent

---

[37]   One might be tempted to even fit an ordered logit model because there seems to be a natural order in the clusters: cluster four provides better growth outcomes than cluster one, cluster one is better than cluster two and cluster two is preferred to cluster three. However, in order to apply the ordered logit model we would have to be sure that there exists a *monotonic* relationship between an underlying index and the observed outcomes (Verbeek, 2004, chap. 7.2.1). Since there is no utility theory underlying our model, we are not certain that this assumption is warranted.

[38]   The smallest number of observations per cluster is 15. It is debateable whether in such a situation a discrete choice model should be applied at all because the estimation method is maximum likelihood and standard errors are valid only asymptotically Verbeek (2004, chap. 7). However, our results are in line with those of the descriptive section, so they appear plausible.

[39]   The results have been derived for a binary logit model only. Given the similar structure of the binary and the multinomial logit model, it can be expected that these results carry over to the multinomial logit model. We are, however, not aware of specific studies concerning this issue.

variables are collected in vector x. Then the following latent variable model can be posted:

$$c_{mk}^* = \gamma_m x_k + \mu_{mk}. \tag{20}$$

The value of the underlying latent variable c_{mk}^* is not directly observable. However, depending on the values of c_{mk}^*, country k is matched to cluster m. If the errors μ are assumed to be independent coming from a Type I extreme value distribution and if all regressors are exogenous, the multinomial logit model is obtained. The growth failure cluster is taken as the reference group in the following. Hence, the probabilities for cluster membership are given by

$$P(C_n = m) = \frac{e^{\gamma_m x_k}}{1 + \sum_{m \neq 3} e^{\gamma_m x_k}} \text{ for m = 1,2,4 and} \tag{21}$$

$$P(C_n = m) = \frac{1}{1 + \sum_{m \neq 3} e^{\gamma_m x_k}} \text{ for m = 3.} \tag{22}$$

Our basic specification uses the percentage of population having completed at least primary education in 1960 to represent education and the GADP indicator to capture the quality of institutions. In order to depict the influence of openness, the openness dummy for 1970 is used. It is tempting to use the number of years a country has been open between 1960 and 2002 as the independent variable. However, since this variable covers the whole sample period and since it is the outcome of policy decisions, it is hard to argue that it is exogenous. Table 15 reports the correlation matrices for several variables separately for each cluster. There are no correlation problems regarding the basic specification in cluster two and cluster three. In cluster four, the correlation between GADP and the average years of schooling in 1960 is verging on 0.6 whereas the percentage of the population having completed at least primary education does not have much correlation with GADP. The most problematic cluster is the first one, where the correlation between GADP and education exceeds 0.65 for both education variables and the correlation between GADP and openness is larger than 0.75. Even though the correlation between GADP and the education using primary school completion rates is somewhat higher than using average years of schooling, we decided to use the former measure because correlation-wise it performs much better in cluster four and not much worse in cluster one. We did not want to use a different institutional quality variable in our basic specification since Jerzmanowski (2006) also relies on GADP. However, given the observed correlations and the small sample size, it would not be too surprising if some typical multicollinearity issues came up in the following.

Columns one in tables 16 and 17 present the results of the basic specification. Whereas table 16 should be used to assess the significance of the variables, table 17 should be referred to in order to learn the direction and the magnitude of the effects that the variables have on the probability to be assigned to a more successful cluster rather than to remain in the growth failure cluster.

Table 15: Correlation Matrices

	GADP	Primary School Completed	Openness Dummy 1970	Investment Rate 1960	Coastal Land	VA 1996	Years liberalized
Cluster 1							
GADP	1						
Primary School Completed	0.790	1.000					
Openness Dummy 1970	0.747	0.649	1.000				
Investment Rate 1960	0.522	0.266	0.354	1.000			
Coastal Land	0.360	0.361	0.662	0.258	1.000		
VA 1996	0.506	0.521	0.517	−0.204	0.333	1.000	
Years liberalized	0.630	0.549	0.960	0.256	0.651	0.602	1.000
Cluster 2							
GADP	1.000						
Primary School Completed	0.175	1.000					
Openness Dummy 1970	−0.288	0.087	1.000				
Investment Rate 1960	0.444	0.622	0.015	1.000			
Coastal Land	0.059	0.460	0.035	0.263	1.000		
VA 1996	0.316	0.707	−0.009	0.442	0.424	1.000	
Years liberalized	−0.006	0.391	0.708	0.261	0.196	0.141	1.000
Cluster 3							
GADP	1.000						
Primary School Completed	0.392	1.000					
Openness Dummy 1970	0.213	0.269	1.000				
Investment Rate 1960	0.018	0.351	0.259	1.000			
Coastal Land	0.275	0.361	0.271	−0.063	1.000		
VA 1996	−0.119	0.289	0.182	0.104	0.407	1.000	
Years liberalized	0.005	0.409	0.712	0.332	0.292	0.483	1.000
Cluster 4							
GADP	1.000						
Primary School Completed	0.126	1.000					
Openness Dummy 1970	−0.186	−0.231	1.000				
Investment Rate 1960	−0.095	−0.751	0.234	1.000			
Coastal Land	−0.156	0.017	−0.385	−0.090	1.000		
VA 1996	0.798	0.098	−0.372	0.101	−0.092	1.000	
Years liberalized	−0.162	−0.272	0.988	0.224	−0.356	−0.376	1.000

GADP is the government antidiversion index, primary school completed denotes the percentage of the population having completed at least primary school, the openness dummy 1970 takes on the value 1 if a country has been open in 1970, coast land stands for the percentage of land within 100km to the coast, VA 1996 is the voice and accountability indicator and years liberalized denotes the number of years a country is characterized as open by Wacziarg and Welch (2008)

Education exerts a positive influence on not ending up as a growth failure. The probability to be assigned to clusters one, two or four increases with additional education. The effect is the largest for moderately successful countries. The average partial effect of a marginal increase in the percentage of population having completed at least primary education is 0.9 percentage points and thus much larger than for clusters one and four. The probability to be assigned to clusters one or four increases noticeably with better institutions. A 0.01 point increase in the quality of institutions as measured

43

Table 16: Multinomial Logit Specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cluster 1							
GADP	8.329*	9.713**	10.300**	10.604**			
	(5.011)	(4.840)	(4.990)	(5.173)			
Primary school completed	0.085**	0.095**	0.111**	0.075*	0.117**	0.114**	0.103**
	(0.041)	(0.040)	(0.045)	(0.041)	(0.040)	(0.041)	(0.041)
Openness Dummy 1970	1.536					1.968	
	(1.335)					(1.272)	
Investment Rate 1960			−0.048				
			(0.055)				
Coastal Land				2.675	3.111*		
				(1.697)	(1.675)		
VA 1996					−0.206	0.085	−0.253
					(0.686)	(0.652)	(0.722)
Years liberalized							0.113**
							(0.049)
Cluster 2							
GADP	0.457	0.480	0.842	1.470			
	(4.060)	(3.957)	(4.062)	(4.348)			
Primary school completed	0.090**	0.092**	0.106**	0.062*	0.061*	0.082**	0.072**
	(0.035)	(0.034)	(0.039)	(0.035)	(0.037)	(0.036)	(0.035)
Openness Dummy 1970	0.307					0.220	
	(1.334)					(1.328)	
Investment Rate 1960			−0.037				
			(0.042)				
Coastal Land				3.705**	3.900**		
				(1.434)	(1.574)		
VA 1996					−0.081	0.398	0.223
					(0.602)	(0.526)	(0.547)
Years liberalized							0.046
							(0.046)
Cluster 4							
GADP	29.429**	31.064**	26.253*	31.679**			
	(14.056)	(13.643)	(13.870)	(14.961)			
Primary school completed	0.158*	0.166**	0.334**	0.167	0.286**	0.234**	0.318*
	(0.086)	(0.084)	(0.168)	(0.104)	(0.115)	(0.085)	(0.168)
Openness Dummy 1970	1.892					3.795	
	(4.385)					(5.533)	
Investment Rate 1960			0.193				
			(0.199)				
Coastal Land				−2.445	−1.349		
				(4.603)	(4.254)		
VA 1996					6.037	6.743*	8.841
					(3.740)	(3.608)	(5.833)
Years liberalized							0.437*
							(0.258)
Pseudo R^2	0.486	0.476	0.494	0.530	0.493	0.471	0.495
χ^2	102.428	100.251	103.982	111.723	103.835	99.158	104.285
p-value	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Standard errors are reported in parenthesis. * denotes significance at the 5%-leve, ** significance at the 1%-level. The number of observations in total is 77. Cluster one contains 17, cluster two 20, cluster three 25 and cluster four 15 observations.

Table 17: Multinomial Logit Specifications - Average Partial Effects

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cluster 1							
GADP	0.308*	0.484**	0.668**	0.585**			
Primary School Completed	0.001**	0.002**	0.000**	0.002*	0.006**	0.004**	0.002**
Openness Dummy 1970	0.154					0.235	
Investment Rate 1960			−0.006				
Coastal Land				0.111	0.120*		
VA 1996					−0.102	−0.133	−0.164
Years liberalized							0.004**
Cluster 2							
GADP	−0.446	−0.562	−0.534	−0.500			
Primary School Completed	0.009**	0.008**	0.009**	0.004*	−0.001*	0.005**	0.004**
Openness Dummy 1970	−0.068					−0.168	
Investment Rate 1960			−0.003				
Coastal Land				0.344**	0.361**		
VA 1996					−0.022	0.035	0.024
Years liberalized							−0.002
Cluster 4							
GADP	0.464**	0.474**	0.312*	0.390**			
Primary School Completed	0.002*	0.002**	0.004**	0.002	0.003**	0.003**	0.004*
Openness Dummy 1970	0.011					0.062	
Investment Rate 1960			0.004				
Coastal Land				−0.092	−0.083		
VA 1996					0.113	0.138*	0.153
Years liberalized							0.006*

In this table the average partial effects of the variables included in the multinomial logit regressions are presented.* denotes significance at the 5%-level, ** significance at the 1%-level. The number of observations in total is 77. Cluster one contains 17, cluster two 20, cluster three 25 and cluster four 15 observations.

by GADP increases the probability to be assigned to cluster one by 0.3 % and that to be assigned to cluster four by 0.46 %. Surprisingly, the openness of the economies is not significant in either of the clusters and a Wald-test on the overall insignificance of this variable cannot be rejected. The result holds regardless of which education indicator is used. Considering the high correlation between GADP and the openness dummy in 1970 in particular in cluster one, the insignificance of this variable occurs most likely because GADP captures already all the variation of the dummy variable. It should not be interpreted as indicating that openness is not important for the cluster assignment. However, as long as the GADP indicator is included in the regression, the openness dummy does not provide additional information. Therefore, columns two report the reduced basic specification, which only uses education and institutional quality as explanatory variables. The Pseudo-R^2 is essentially unchanged. The only noticeable change is the increased influence of the quality of institutions on the probability to be assigned to cluster one. It seems now that the indicator of openness is removed from the equation, the GADP indicator captures even more thoroughly the influence of being liberalized.

Given the reduced basic specification we next analyze to what extent the results are affected by introducing additional explanatory variables such as the investment propensity or variables referring to the structural features of the clusters. One of the results is given in columns three in tables 16 and 17, where the investment propensity is included as an additional explanatory variable. The investment propensity is not significant in any of the clusters and does not change the significance of the basic variables. A Wald test on the overall significance of the investment propensity corroborates that the investment propensity has no influence on the predicted cluster probabilities. It is somewhat concerning that the average partial effect of the GADP indicator in cluster one is considerably larger than before. The sensitivity of the estimator points at a potential multicollinearity problem and indeed the correlation between GADP and the investment propensity in cluster one exceeds 0.5. Additional structural variables have been introduced, but except for the percentage of land or population within one hundred kilometers to the coast all other variable can safely be discarded. The results including the land variable are presented in columns four. The land variable is highly significant in determining whether a country belongs to the moderately successful cluster two rather than to the growth failure cluster three. Besides education, geographic conditions appear to play an important role in shaping a positive environment for growth. The results for cluster one do not change significantly. In cluster four the educational variable becomes insignificant at the five percent level, even though none of the included variables are highly correlated with each other.

So far, the quality of institutions appears to exert a robust effect on the probabilities to belong to the successful growth clusters one and four. This, however, is a special feature of the GADP indicator. If the voice and accountability indicator is chosen instead, the significance of the quality

of institutions vanishes (see columns five). The reason for the stronger performance of the GADP indicator lies in what it exactly measures. The GADP indicator emphasizes both property rights and contractual reliability, which have been singled out as the most important institutional features of economies (Acemoglu, 2009, chap.4.5). If the voice and accountability indicator is used, these aspects are only indirectly captured through their correlation with freedom. Unlike in the previous specification, the land variable is now significantly related to the probability of being assigned to cluster one. Since the GADP indicator and the land variable are both highly correlated with the openness of the economies and the years of liberalized trade in this cluster, it is conceivable that the significant land variable captures the aspect of openness, which is not as fully captured by the voice and accountability indicator as it is by the GADP indicator (see table 15). In order to shed further light on the importance of openness, we reestimate our basic specification using the voice and accountability indicator. If the land variable indeed captures openness, it could be the case that the openness dummy becomes significant with the new institutional quality variable. However, columns six show that this is not the case. Openness only becomes a significant predictor if the years of liberalized trade are introduced as an explanatory variable (see columns seven). Since this variable is most likely endogenous, it is not clear what kind of bias is implicit in the estimates. Yet, if one looks at the magnitude of the institutional and educational effects, these are not drastically changed compared to the regression with the openness dummy. We interpret this fact as weak evidence that the bias due to endogeneity is not enormous and that openness of the economies should not be rejected as a possible determinant of the cluster assignment.

Summarizing the previous results, the quality of institutions as measured by the GADP index is a very useful indicator to predict the association of countries with either cluster one or cluster four. It is, however, not very useful in order to distinguish between growth failures and moderately successful countries. For these two groups, a combination of the geographic conditions and the efforts in education is more useful. The influence of openness on the assignment of countries to clusters remains unclear: on the one hand, there is always one significant variable that is highly correlated with openness, so that openness could be indirectly driving the results. We conjecture that in particular the geographic variables indirectly capture the effects of openness, since certain geographic conditions make openness easier and more worthwhile to achieve. On the other hand,

Table 18: Multinomial Logit Specifications - Share Correctly Predicted

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Cluster 1	0.611	0.667	0.667	0.722	0.444	0.500	0.556
Cluster 2	0.348	0.304	0.348	0.522	0.348	0.391	0.435
Cluster 3	0.750	0.750	0.786	0.714	0.714	0.714	0.750
Cluster 4	0.933	0.933	0.933	0.933	0.933	0.933	0.933

only the years of liberalized trade succeed in achieving a significant influence of openness on the assignment to clusters. In order to understand more thoroughly the importance of the different variables, we analyze the success of the model specifications to correctly predict the assignment of the sample countries to the clusters (see table 18). The basic reduced model specification is able to predict very well countries belonging to cluster four and reasonably well the countries belonging to cluster one. In cluster one, mainly countries that are "late developers" like China or Brazil cannot be identified. This indicates that policy decisions also in later stages can indeed determine the dynamics of growth. However, our model is not able to capture these, because we focus on variables at the beginning of the sample period only. We checked to what extent the correct prediction of countries belonging to clusters one and four are driven by the GADP variable. The percentage of countries correctly assigned to clusters one and four would remain unchanged. The basic model specification is not very successful in correctly predicting membership in cluster two: less than 50 percent of the countries are correctly classified. The countries are misclassified into cluster three and cluster one, whereby the misclassification into cluster three dominantes. Only model specification (4) or (7), i. e. the ones that include either a geographic variable or the years of liberalized trade, succeed noticeably better in identifying the countries of cluster. The identification of countries belonging to cluster one deteriorates considerably if a different institutional variable to GADP is used.

In summary, the preceding descriptive and quantitative analyzes have revealed that the government antidiversion index is indeed well suited to single out countries that have achieved successful growth dynamics during the sample period. It appears that the driving factor of being really successful are institutions that guarantee a large amount of certainty in conducting business. If uncertainty prevails countries are generally either moderately successful growers or growth failures. The GADP index does not help to distinguish between these two categories of countries. Rather, if institutions are bad, countries have to draw on other factors in order to ensure reasonable growth. Education has been identified as one of these factors. Apart from that, geographic variables are of importance. We suspect that these geographic variables capture to some extent the integration into the world economy, but this could not be formally proven. However, even these additional factors are hardly sufficient to differentiate between clusters two and three. It appears that there are other, less easily measured factors that are of importance. In light of these results, conditioning the transition probabilities of the Markov switching process on the GADP index only is a somewhat too narrow interpretation. Such a specification misses the mitigating effects that a well educated workforce or an advantageous geographic location can have in the presence of weak institutions.

6 Conclusion

The within country variation of growth rates has been identified as an important research topic. This paper contributes to the discussion by interpreting the process of growth within a Markov switching model. The growth pattern of each country is determined by the available growth regimes and by the transition probabilities that determine how often and how long a country visits each growth regime. In order to allow for countries to have varying transition probabilities without pre-determining the variables that are allowed to influence these probabilities, the Markov switching approach is enhanced by a clustering mechanism. The model is estimated for 84 countries during the period from 1962 to 2002. In the estimation process the growth regimes and the transition probabilities are determined endogenously. Each country is assigned to the cluster which maximizes the likelihood of observing the actual growth data.

The estimated model identifies four growth regimes and three clusters. The growth regimes are similar to those obtained by Jerzmanowski (2006). There is a stable growth regime in which countries grow steadily with a long-run growth rate of around two percent, and there is a miracle growth state in which countries embark on a catch-up process. There is also a stagnation regime, which is characterized by zero growth on average and relatively unsteady growth. Finally, a crisis state is identified that captures very large shocks, both positive and negative ones, although on average negative shocks are more frequent. The identified clusters imply growth processes with different characteristics. Countries in the first cluster are characterized by steady growth and pro-longed growth spurts. The list of countries is made up by industrialized and newly industrialized countries as well as some steadily growing developing countries. The second and third cluster consist mainly of developing countries. Countries in the second cluster spend the majority of their time in the stagnation and the stable growth regime with approximately equal weight. Growth spurts do occur, but they are short-lived compared to those of cluster one. Finally, countries in the third cluster usually find themselves either in stagnation and crisis, which implies a very unsteady behavior of the growth rates. The average growth of the countries is noticeably lower than that of industrial leader countries such as the United States so that countries in this cluster fall relatively more behind. Not surprisingly, the majority of African countries is associated with this cluster.

In a final step the thus defined growth clusters are used to gain some knowledge about the growth fundamentals that are responsible for the different transition matrices. The focus is on identifying fundamentals that have enabled the initially non-industrialized countries to successfully embark on a development process. Three broad categories of fundamentals are considered: factors pertaining to human and physical capital accumulation, factors relating to important structural features of the economies and the quality of institutions. A descriptive and quantitative analysis reveals that the most successful countries are characterized by reasonably good institutions, moderate amounts

of human capital in the beginning of the sample period and to some extent favorable geographic conditions such as easy access to the sea. Compared to the moderately successful countries in cluster two, cluster one countries are distinguished by their better quality of institutions. Human capital and geographic conditions seem to be particularly important in preventing countries from becoming growth failures. Since geographic conditions and openness of the economies is correlated, the importance of geographic conditions might in reality indicate the importance of opening up the economies in order to learn by importing and exporting goods and services. With regard to the appropriate conditioning variable for the transition probabilities in a Markov switching model, these findings indicate that one conditioning variable may not be enough. The quality of institutions is a good indicator to capture the growth dynamics of successful countries. It is less suitable to distinguish between moderately successful countries and growth failures. Therefore, conditioning the transition probability matrix on the quality of institutions only may in this case falsely attribute too positive a transition matrix to a growth failure country and vice versa. A more comprehensive collection of conditioning variables is required.

The following policy conclusions emerge: it is of utmost importance for developing countries to enhance their human capital base, because an educated workforce seems to be a suitable means to mitigate negative effects for instance from bad institutions. This should be the first step in any development plan, partly because of its importance and partly because it is a measure relatively easy to implement without too much resistance from vested interest groups. Countries with favorable geographic conditions seem to have an inherent advantage in succeeding to reach a successful or moderately successful growth path. The most likely reason for this advantage are the lower transportation costs that make it easier and more worthwhile for these countries to trade or to attract foreign direct investment. Hence, development policies should also focus on the expansion of road, railway and water networks in order to facilitate transportation and trading. Eventually, however, any development plan has to tackle the difficult issue of institutional reform, because good institutions appear to be a prerequisite for sustained growth. Yet, this measure is likely to be the most difficult to implement.

While this paper has identified clusters of countries the growth process of which can be described by the same transition probability matrix, it has abstracted from one potentially very important aspect: the analysis does not allow countries to undergo structural change and as a result switch into a cluster defined by a more favorable transition matrix. However, if one advises developing countries to invest in human capital, to open up their economies or to work on their institutions, it is precisely this effect that one hopes for. As soon as enough data is available the Markov switching growth model should be extended into this direction. For the moment, a first valuable step is to know what made countries achieve favorable transition matrices in the past.

References

Acemoglu, D. (2009). *Introduction to Modern Economic Growth.* Princeton University Press.

Acemoglu, D., S. Johnson, and J. Robinson (2003). Institutional Causes, Macroeconomic Symptoms: Volatility, Crises and Growth. *Journal of Monetary Economics 50,* 49–123.

Acemoglu, D., S. Johnson, and J. A. Robinson (2001). The Colonial Origins of Comparative Development: An Empirical Investigation. *American Economic Review 91,* 1369–1401.

Aizenman, J. and M. Spiegel (2007). Takeoffs. (13084).

Alfo, M., G. Trovato, and R. Waldmann (2008). Testing for Country Heterogeneity in Growth Models Using a Finite Mixture Approach. *Journal of Applied Econometrics 23,* 487–514.

Alon, J., S. Sclaroff, and G. Kollios (2003). Discovering Clusters in Motion Time-Series Data. *Computer Vision and Pattern Recognition. Proceedings. 2003 IEEE Computer Society Conference on 1,* 375–381.

Ardic, O. P. (2006). The Gap Between the Rich and the Poor: Patterns of Heterogeneity in the Cross-Country Data. *Economic Modelling 23,* 238–555.

Asea, P. K. and B. Blomberg (1998). Lending Cycles. *Journal of Econometrics 83,* 89–128.

Azariadis, C. and A. Drazen (1990). Threshold Externalities in Economic Development. *Quarterly Journal of Economics 105,* 501–526.

Barro, R. (2000). International Data on Educational Attainment: Updates and Implications. (7911).

Barro, R. J. and J.-W. Lee (2001). International data on educational attainment: updates and implications. *Oxford Economic Papers 3,* 541–563.

Basturk, N., R. Paap, and D. van Dijk (2008). Structural Differences in Economic Growth. (2008-085/4).

Becker, T. and P. Mauro (2006). Output Drops, and the Shocks that Matter.

Berg, A., C. Leite, and J. D. a. Ostry (2008). What Makes Growth Sustained?

Berthelemy, J.-C. (2006). Convergence and Development Traps: How Did Emerging Economies Escape the Underdevelopment Trap?

Biernacki, C., G. Celeux, and G. Govaert (2003). Choosing Starting Values for the EM Algorithm for Getting the Highest Likelihood in Multivariate Gaussian Mixture Models. *Computational Statistics & Data Analysis 41,* 561–575.

Bloom, D. E., D. Canning, and J. Sevilla (2003). Geography and Poverty Traps. *Journal of Economic Growth 8*, 335–378.

Brock, W. A. and S. N. Durlauf (2001). Growth Empirics and Reality. *The World Bank Economic Review 15*, 229–272.

Bryant, P. G. (1991). Large-Sample Results for Optimization-Based Clustering Methods. *Journal of Classification 8*, 31–44.

Butler, M. (2003). *Hidden Markov Model Clustering of Acoustic Data.* Online: http://www.inf.ed.ac.uk/publications/thesis/online/IM030057.pdf.

Cadez, I. and Heckermann (2003). Model-Based Clustering and Visualization of Navigation Patterns on a Web Site. *Data Mining and Knowledge Discovery 7*, 399–424.

Campbell, S. (2002). Specification Testing and Semiparametric Estimation of Regime Switching Models: An Examination of the US Short Term Interest Rates. (2002-26).

Canova, F. (2004). Testing For Convergence Clubs in Income Per Capita: A Predictive Density Approach. *International Economic Review 45*, 49–77.

Celeux, G. and G. Govaert (1993). Comparison of the Mixture and the Classification Maximum likelihood in Cluster Analysis. *Journal of Statistical Computation and Simulation 47*, 127–146.

Chen, S.-W. (2007a). Using Regional Cycles to Measure National Business Cycles in the U.S. with the Markov Switching Panel Model. *Economic Bulletin 3*, 1–12.

Chen, S.-W. (2007b). Identifying US turning points revisited: the panel model with the regime switching approach. *Applied Economics Letters*, 1–5.

Cohen, D. and M. Soto (2001). Growth and Human Capital: Good Data, Good Results. (3025).

Collier, P. (2006). African Growth: Why a 'Big Push'? *Journal of African Economies 15*, 188–211.

Cramer, J. S. (2007). Omitted Variables and Misspecified Disturbances in the Logit Model. *Oxford Bulletin of Economics and Statistics 69*, 545–555.

Dasgupta, S. (1998). Patterns of Trade and Growth Under Increasing Returns: Escapte From The Poverty Trap - A Comment. *The Japanese Economic Review 49*, 234–247.

Dempster, A., N. Laird, and D. Rubin (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B 39*, 1–38.

Desdoigts, A. (1999). Patterns of Economic Development and the Formation of Clubs. *Journal of Economic Growth 4*, 305–330.

Diebold, F. X., L. Joon-Haeng, and G. C. Weinbach (1994). Regime switching with time-varying transition probabilities. pp. 283–302.

Dunmur, A. P. and D. M. Titterington (1998). The Influence of Initial Conditions on Maximum Likelihood Estimation of the Parameters of a Binary Hidden Markov Model. *Statistics & Probability Letters 1998*, 67–73.

Durlauf, S. N. (2001). Econometric analysis and the study of economic growth: a sceptical perspective. *Macroeconomics and the real World 1*, 249–261.

Durlauf, S. N. and P. A. Johnson (1995). Multiple Regimes and Cross-Country Growth Behaviour. *Journal of Applied Econometrics 10*, 365–384.

Easterly, W. and R. Levine (1997). Africa's Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics 112*, 1203–1250.

Fruehwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models*. Springer.

Gallup, J. L., J. D. Sachs, and A. D. Mellinger (1999). Geography and Economic Development. *International Regional Science Review 22*, 179–232.

Garcia, R. (1998). Asymptotic Null Distribution of the Likelihood Ratio Test in Markov Switching Models. *International Economic Review 39*, 763–788.

Glaeser, E. L., R. La Porta, and F. Lopez-de Silanes (2004). Do Institutions Cause Growth? *Journal of Economic Growth 9*, 271–303.

Hall, R. and C. I. Jones (1999). Why Do Some Countries Produce So Much More Output per Worker than Others? *Quarterly Journal of Economics 114*, 83–116.

Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica 57*, 357–384.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics 45*, 39–70.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hansen, B. E. (1992). The Likelihood Ratio Test under Non-Standard Conditions: Testing the Markov Switching Model of GNP. *Journal of Applied Econometrics 7*, S61 – S82.

Hansen, B. E. (2000). Sample Splitting and Threshold Estimation. *Econometrica 68*, 575–603.

Hausmann, R., L. Pritchett, and D. Rodrik (2005). Growth Accelerations. *Journal of Economic Growth 10*, 303–329.

Hausmann, R., F. Rodriguez, and R. Wagner (2006). Growth Collapses. (RWP06-046).

Hausmann, R., D. Rodrik, and A. Velasco (2004). Growth Diagnostics.

Heston, A., R. Summers, and B. Aten (2006). Penn World Table Version 6.2.

Imbens, G. W. and J. M. Wooldridge (2007). Control Function and Related Methods.

Jedidi, K., S. J. Harsharanjeet, and W. S. DeSarbo (1997). Finite-Mixture Structural Equation Models for Response-Based Segmentation and Unobserved Heterogeneity. *Marketing Science 16*, 39–59.

Jerzmanowski, M. (2006). Empirics of Hills, Plateaus, Mountains and Plains: A Markov-switching Approach to Growth. *Journal of Development Economics 81*, 357–385.

Jones, B. F. and B. A. Olken (2008). The Anatomy of Start-Stop Growth. *Review of Economics and Statistics 90*, 582–587.

Jong-A-Pin, R. and J. de Haan (2007). Political Regime Change, Economic Reform and Growth Accelerations. (1905).

Karlis, D. and E. Xekalaki (2003). Choosing Initial Values for the EM Algorithm for Finite Mixtures. *Computational Statistics & Data Analysis 41*, 577–590.

Kaufmann, D., A. Kraay, and M. Mastruzzi (2008). Governance Matters VII: Aggregate and Individual Governance Indicators 1996-2007. (4654).

Kim, C.-J. (2004a). Markov-switching models with endogenous explanatory variables. *Journal of Econometrics 122*, 127–136.

Kim, C.-J. (2004b). Markov-Switching Models with Endogenous Explanatory Variables II: A Two-Step MLE Procedure with Standard-Error Correction.

Kim, C.-J. and C. R. Nelson (1999). *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications.* MIT Press.

Kim, C.-J., J. Piger, and R. Startz (2008). Estimation of Markov regime-switching regression models with endogenous switching. *Journal of Econometrics 143*, 263–273.

King, R. G. and R. Levine (1993a). Finance and Growth: Schumpeter Might be Right. *Quarterly Journal of Economics 108*, 717–737.

King, R. G. and R. Levine (1993b). Finance, Entrepreneurship, and Growth: Theory and Evidence. *Journal of Monetary Economics 32*, 513–542.

Knab, B. (2000). *Erweiterungen von Hidden-Markov-Modellen zur Analyse ökonomischer Zeitreihen.*

Knack, S. and P. Keefer (1995). Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures. *Economics And Politics 7*, 207–227.

Kose, M. A., C. Otrok, and C. H. Whiteman (2003). International Business Cycles: World, Region, and Country-Specific Factors. *American Economic Review 93*, 1216–1239.

Kourtellos, A. (2002). A Projection Pursuit Appraoch to Cross Country Growth Data. (2002-13).

Krolzig, H.-M. (1997). *Markov Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*. Springer.

Li, C. and G. Biswas (2000). A Bayesian Approach to Temporal Data Clustering using Hidden Markov Models.

Lutz, W., A. Goujon, and K. C. Samir (2007). Reconstruction of Populations by Age, Sex and Level of Educational Attainment for 120 Countries for 1970-2000.

Majumdar, M. and T. Mitra (1995). Patterns of Trade and Growth Under Increasing Returns: Escape From the Poverty Trap. *The Japanese Economic Review 46*, 207–223.

McLachlan, G. J. (1982). The Classification and Mixture Maximum Likelihood Approaches to Cluster Analysis. pp. 199–208.

McLachlan, G. J. and T. Krishnan (1997). *The EM Algorithm and Extensions*. John Wiley & Sons.

Nelson, R. R. (1956). A Theory of the Low-Level Equilibrium Trap in Underdeveloped Economies. *American Economic Review 45*, 894–908.

Paap, R., P. H. Franses, and D. van Dijk (2005). Does Africa Grow Slower than Asia, Latin America and the Middle East? Evidence from a New Data-Based Classification method. *Journal of Development Economics 77*, 553–570.

Papageorgiou, C. (2002). Trade as a Threshold Variable for Multiple Regimes. *Economics Letters 77*, 85–91.

Pritchett, L. (2000). Understanding Patterns of Economic Growth: Searching for Hills among Plateaus, Mountains, and Plains. *The World Bank Economic Review 14*, 221–250.

Pritchett, L. (2003). A Toy Collection, a Socialist Star and a Democratic Dud? See Rodrik (2003), pp. 123–151.

Psaradakis, Z. and M. Sola (1998). Finite-Sample Properties of the Maximum Likelihood Estimator in Autoregressive Models with Markov Switching. *Journal of Econometrics 86*, 369–386.

Ramalho, E. A. and J. J. S. Ramalho (2007). Is Neglected Heterogeneity Really an Issue in Nonlinear Models? A Simulation Exercise for Binary and Fractional Data. (2007/08).

Rodrik, D. (2003). *In Search of Prosperity: Analytic Narratives on Economic Growth.* Princton University Press.

Shaffer, J. P. (1995). Multiple Hypothesis Testing. *Annual Review of Psychology 46*, 561–584.

Smith, A., P. A. Naik, and C.-L. Tsai (2006). Markov-switching model selection using Kullback-Leibler divergence. *Journal of Econometrics 134*, 553–577.

Temple, J. (1999). The New Growth Evidence. *Journal of Economic Literature 37*, 112–156.

Verbeek, M. (2004). *A Guide to Modern Econometrics.* John Wiley & Sons, Ltd.

Wacziarg, R. and K. H. Welch (2008). Trade Liberalization and Growth: New Evidence. *The World Bank Economic Review 22*, 187–231.

Wichern, B. (2001). *Hidden-Markov-Modelle zur Analyse und Simulation von Finanzmarktreihen.* Hundt Druck GmbH,.

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

WorldBank (2007). *World Development Indicators.* Digitale Bibliothek, FU Berlin.

Appendix A

In this section the derivation of equation (7) is explained in more detail. Collect all parameters characterizing equation 5 in vector θ^1, the transitions probabilities in vector θ^2 and collect θ^1 and θ^2 in θ. The density of an observation (x_t, y_t) conditional on the state taking on the value j and conditional on the parameters in θ is given by

$$f(y_t|s_t = j, x_t, \theta) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(\frac{(y_t - x_t'\beta_j)^2}{2\sigma_j^2}\right). \tag{23}$$

Define $\mathcal{Y}_t = (y_1, y_2, \ldots, y_t)$, $\mathcal{X}_t = (x_1, x_2, \ldots, x_t)$ and $\mathcal{S}_t = (s_1, s_2, \ldots, s_t)$ as in the main text. The complete-data likelihood function is the product of all (y_t, x_t) augmented by the state s_t, which is supposed to be known with certainty, i. e.

$$
\begin{aligned}
\mathcal{L}_T &= f(\mathcal{Y}_T, \mathcal{S}_T|\mathcal{X}_T, \theta) \\
&= f(y_1, s_1|x_1, \theta) \prod_{t=2}^{T} f(y_t, s_t|\mathcal{Y}_{t-1}, \mathcal{S}_{t-1}, \mathcal{X}_{t-1}, x_t, \theta).
\end{aligned}
$$

This expression can be simplified by noting the following:

- The joint density of y_t and s_t can be expressed as the product of the conditional and marginal probabilities, i. e.

$$f(y_t, s_t|\mathcal{Y}_{t-1}, \mathcal{S}_{t-1}, \mathcal{X}_{t-1}, x_t, \theta) = f(y_t|s_t, \mathcal{Y}_{t-1}, \mathcal{S}_{t-1}, \mathcal{X}_{t-1}, x_t, \theta) P(s_t|\mathcal{Y}_{t-1}, \mathcal{S}_{t-1}, \mathcal{X}_{t-1}, x_t, \theta).$$

- Since the states follow a first-order Markov process, $P(s_t|\mathcal{Y}_{t-1}, \mathcal{S}_{t-1}, \mathcal{X}_T, \theta)$ does not depend on the whole history of states, but only on s_{t-1}. Moreover, the states s_t in the model are independent of $\mathcal{X}_t$ and $\mathcal{Y}_t$. Hence,

$$P(s_t|\mathcal{Y}_{t-1}, \mathcal{S}_{t-1}, \mathcal{X}_{t-1}, x_t, \theta) = P(s_t|s_{t-1}, \theta).$$

- The conditional density of y_t does not depend on the history of states $\mathcal{S}_{t-1}$. Hence this information can be dropped.

- Summarize the history of the independent and dependent variable up to period $t-1$ by $\psi_{t-1} = \{\mathcal{Y}_{t-1}, \mathcal{X}_{t-1}\}$.

Using all four simplifications the complete-data likelihood function can be expressed as

$$
\begin{aligned}
\mathscr{L}_T &= f(y_1, s_1|\psi_0, x_1, \theta) \prod_{t=2}^{T} f(y_t, s_t|\psi_{t-1}, x_t, \theta) \\
&= f(y_1|s_1, \psi_0, x_1, \theta)P(s_1|\psi_0, \theta) \prod_{t=2}^{T} f(y_t|s_t, \psi_{t-1}, x_t, \theta)P(s_t|s_{t-1}, \theta) \\
&= f(y_1|s_1, \psi_0, x_1, \theta^1)P(s_1|\psi_0, \theta^2) \prod_{t=2}^{T} f(y_t|s_t, \psi_{t-1}, x_t, \theta^1)P(s_t|s_{t-1}, \theta^2)
\end{aligned}
$$

Strictly speaking, $f(y_t|s_t, \psi_{t-1}, x_t, \theta^1)$ could be further simplified to $f(y_t|s_t, x_t, \theta^1)$ because in our specification the history of observations is not relevant for the conditional density of the current observation. Nevertheless, ψ_{t-1} will be retained in the formulation in order to clarify at later stages whether the information up to period t or whether the information of the whole sample is used to infer state probabilities.

In order to arrive at equation (7) the indicator function as defined in the main text has to be introduced. Effectively, each single observation is thus expressed as a sum of all possible contingencies how the observation could have been generated. The indicator function ensures that only the true contingency is considered in the following calculations. Notice further that $P(s_t = j|s_{t-1} = i, \theta^2)$ is equivalent to p_{ij}, since the transition probabilities do not depend on other variables than the defined parameters. Moreover, $\rho_j = P(s_0 = j|\psi_0, \theta^2)$ holds. The complete-data likelihood function therefore can be reformulated as follows:

$$
\begin{aligned}
\mathscr{L}_T &= \left[\sum_{j=1}^{Q} I(s_1 = j)f(y_1|s_1 = j, \psi_0, x_1, \theta^1)\rho_j \right] \\
&\times \prod_{t=2}^{T} \left\{ \sum_{j=1}^{Q} \sum_{i=1}^{Q} I(s_t = j|s_{t-1} = i)f(y_t|s_t = j, \psi_{t-1}, x_t, \theta^1)p_{ij} \right\} \quad (24)
\end{aligned}
$$

Taking logs of (24) and rearranging gives equation (7) in the main text. By separating θ into θ^1 and θ^2, it is immediately obvious that derivatives of the log-likelihood function with regards to the two parts of θ are independent of each other, which facilitates the computational process considerably.[40]

Appendix B

This appendix shortly points out why the chosen algorithm succeeds in finding a local maximum. The suitability of the algorithm has been shown by Knab (2000, chap. 4.2.1) for the case that the

---

[40] Just as the main section, this appendix is based on on Hamilton (1994, chap. 22), Krolzig (1997, chap. 5 and 6), Kim and Nelson (1999, chap. 4) and Diebold et al. (1994).

model clusters differ both with respect to the states and to the transition probabilities. We will first introduce the proof that the proposed algorithm increases the the value of the log-likelihood function in each iteration step before arguing that the same line of argument applies for the model presented in this paper. The monotonicity of the usual EM algorithm, i. e. the fact that the value of the objective function increases in each iteration step until a stationary point is reached, is presupposed. For a detailed proof thereof, cf. for instance McLachlan and Krishnan (1997).

Consider first a Markov switching model with $m = 1, \ldots, M$ clusters, where each cluster features different states and transition probabilities. The variables are defined in the same way as in the main text. Suppose you have completed the $iter^{\text{th}}$ iteration step of the Markov switching classification algorithm presented in section 3.2. In this case the expected log-likelihood function and its evolution during one algorithm step is given by

$$= \sum_{k=1}^{K} \sum_{m=1}^{M} C_m^{iter} \tilde{\ell}_{ck}(\mathcal{O}^k | \theta_m^{iter}) \tag{25}$$

$$\leqslant \sum_{k=1}^{K} \sum_{m=1}^{M} \max_{m=1,\ldots,M} \tilde{\ell}_{ck}(\mathcal{O}^k | \theta_m^{iter}) \tag{26}$$

$$= \sum_{k=1}^{K} \sum_{m=1}^{M} C_m^{iter+1} \tilde{\ell}_{ck}(\mathcal{O}^k | \theta_m^{iter}) \tag{27}$$

$$\tag{28}$$

$$\leqslant \sum_{k=1}^{K} \sum_{m=1}^{M} C_m^{iter+1} \tilde{\ell}_{ck}(\mathcal{O}^k | \theta_m^{iter+1}) \tag{29}$$

Equation (25) corresponds to the state of the algorithm after the expectation step given the current classification of time-series to clusters. The value of the incomplete data log-likelihood function may be improved if there are countries for which the $\tilde{\ell}_{ck}$ in another cluster than the one currently assigned to exceeds that of the current cluster (equation (26)). In the classification step all countries are assigned to the cluster that maximizes $\tilde{\ell}_{ck}$ (equation (27)). Finally, given this new cluster assignment the maximization step of the EM algorithm is carried out. Since the assignment to clusters remains unchanged in this step, the monotonicity of the EM algorithm ensures that the last inequality holds for the incomplete data log-likelihood function as a whole.

The model in this paper effectively corresponds to the model used by Knab (2000) with a set of linear restrictions imposed on the parameters. Therefore, it has to be established which steps of the presented proof are sensitive to the introduction of the restriction. Obviously, equations (25) - (27) continue to hold because these steps require comparisons of the different $\tilde{\ell}_{ck}$ values for one country only. The crucial question therefore becomes whether the EM algorithm maintains its monotonicity when linear parameter restriction are involved. **?** prove that this is the case. Hence, it follows that (29) continues to apply in the restricted model: the value of the incomplete data

log-likelihood function is never decreased in an iteration step. Fortunately, the imposed restrictions are easy to implement and a closed form solution can be obtained in the maximization step. Thus, we do not have to recur to the numerical methods suggested in **?**.

Appendix C

Table 19: Average Regime Probabilities

Country	Cluster	Crisis	Stagnation	Stable Growth	Miracle Growth	Average Growth Rate
Algeria	1	0.18	0.48	0.26	0.08	1.33
Argentina	1	0.14	0.69	0.10	0.07	0.58
Australia	2	0.00	0.01	0.99	0.00	2.43
Austria	2	0.00	0.10	0.73	0.18	2.88
Belgium	2	0.00	0.03	0.91	0.05	2.77
Benin	1	0.09	0.40	0.45	0.05	0.89
Bolivia	3	0.03	0.23	0.63	0.11	0.50
Brazil	2	0.00	0.22	0.39	0.39	2.28
Burkina Faso	3	0.01	0.65	0.15	0.19	1.24
Burundi	1	0.22	0.70	0.02	0.06	0.74
Cameroon	1	0.11	0.49	0.34	0.05	0.85
Canada	2	0.00	0.02	0.97	0.01	2.45
Chad	1	0.25	0.70	0.02	0.03	-0.68
Chile	3	0.05	0.35	0.41	0.20	2.18
China	2	0.00	0.16	0.00	0.84	6.69
Colombia	2	0.00	0.00	1.00	0.00	1.89
Costa Rica	3	0.01	0.23	0.63	0.13	1.77
Cote d'Ivoire	1	0.16	0.62	0.17	0.06	1.00
Denmark	3	0.00	0.16	0.71	0.12	2.12
Dominican Republic	3	0.03	0.32	0.49	0.17	3.16
Ecuador	1	0.07	0.35	0.54	0.05	1.52
Egypt	3	0.03	0.33	0.48	0.16	2.89
El Salvador	2	0.00	0.07	0.92	0.00	1.18
Ethiopia	1	0.16	0.39	0.44	0.02	1.32
Finland	3	0.01	0.28	0.54	0.17	2.66
France	2	0.00	0.00	0.99	0.00	2.69
Ghana	1	0.36	0.53	0.09	0.03	2.74
Greece	2	0.00	0.08	0.61	0.30	3.11
Guatemala	2	0.00	0.01	0.99	0.00	0.98
Guinea	3	0.02	0.62	0.19	0.17	-0.27
Honduras	3	0.02	0.55	0.26	0.18	0.75

Country	Cluster	Crisis	Stagnation	Stable Growth	Miracle Growth	Average Growth Rate
Hong Kong	2	0.01	0.23	0.11	0.65	5.37
India	3	0.01	0.32	0.49	0.18	3.01
Indonesia	3	0.02	0.44	0.32	0.22	3.36
Iran	1	0.23	0.53	0.07	0.17	1.72
Ireland	2	0.00	0.05	0.64	0.31	4.14
Israel	3	0.03	0.39	0.41	0.17	2.62
Italy	2	0.00	0.08	0.78	0.13	2.75
Jamaica	3	0.01	0.46	0.40	0.13	0.61
Japan	2	0.00	0.05	0.70	0.26	3.97
Kenya	3	0.01	0.41	0.45	0.13	0.21
South Korea	2	0.00	0.13	0.03	0.84	6.35
Madagascar	1	0.08	0.59	0.31	0.01	-1.28
Malawi	1	0.15	0.48	0.31	0.06	1.29
Malaysia	2	0.02	0.16	0.23	0.58	4.77
Mali	1	0.19	0.64	0.12	0.05	1.10
Mexico	3	0.01	0.29	0.53	0.16	1.92
Morocco	3	0.01	0.50	0.29	0.20	2.50
Mozambique	1	0.16	0.52	0.18	0.13	1.38
Nepal	3	0.01	0.42	0.40	0.17	1.52
Netherlands	2	0.00	0.01	0.99	0.01	2.33
New Zealand	2	0.00	0.17	0.66	0.17	1.45
Nicaragua	1	0.13	0.62	0.22	0.04	-0.73
Niger	1	0.18	0.69	0.06	0.07	-0.92
Nigeria	1	0.21	0.59	0.14	0.06	0.43
Norway	2	0.00	0.00	1.00	0.00	3.11
Pakistan	3	0.01	0.16	0.70	0.13	2.92
Panama	3	0.03	0.30	0.50	0.17	2.83
Paraguay	2	0.00	0.08	0.78	0.14	1.48
Peru	1	0.12	0.46	0.37	0.05	0.71
Philippines	3	0.01	0.35	0.48	0.15	1.36
Portugal	2	0.00	0.12	0.57	0.31	3.79
Rwanda	1	0.29	0.56	0.04	0.12	0.75
Senegal	1	0.15	0.72	0.09	0.05	-0.48
Singapore	2	0.00	0.23	0.02	0.75	4.72
South Africa	2	0.00	0.00	1.00	0.00	1.33
Spain	2	0.00	0.05	0.72	0.23	3.37
Sri Lanka	3	0.01	0.26	0.52	0.21	3.83
Sweden	2	0.00	0.00	0.99	0.00	2.05
Switzerland	2	0.00	0.05	0.93	0.02	1.44
Syria	1	0.34	0.43	0.14	0.09	2.19
Taiwan	2	0.00	0.10	0.03	0.87	6.67

Country	Cluster	Crisis	Stagnation	Stable Growth	Miracle Growth	Average Growth Rate
Tanzania	1	0.10	0.61	0.28	0.02	1.77
Thailand	2	0.00	0.19	0.04	0.77	4.88
Togo	1	0.15	0.70	0.07	0.08	-0.40
Tunisia	3	0.01	0.36	0.45	0.18	3.26
Turkey	3	0.01	0.46	0.32	0.21	2.29
Uganda	1	0.12	0.47	0.37	0.04	0.76
United Kingdom	2	0.00	0.01	0.99	0.00	2.31
United States	2	0.00	0.01	0.98	0.01	2.51
Uruguay	3	0.03	0.61	0.16	0.19	0.92
Venezuela	1	0.14	0.69	0.05	0.11	0.07
Zambia	1	0.23	0.66	0.06	0.04	0.03
Zimbabwe	1	0.29	0.56	0.07	0.09	0.08

In this table, the average smoothed state probabilities, i. e. the probability for a country to be in a certain regime in an average year are reported. Formally, each column equals $(1/T)\sum_{t=1}^{T} P(s_t = j|\psi_T)$, where $P(\cdot)$ is the estimated probability of country k being in state j given the information of the entire sample and conditional on its cluster assignment.

Diskussionsbeiträge
des Fachbereichs Wirtschaftswissenschaft
der Freien Universität Berlin

2009

2009/1   ENGLER, Philipp
            Global Rebalancing in a Three-Country Model
            Volkswirtschaftliche Reihe

2009/2   MUCHLINSKI, Elke
            Is there a need for a coded language in central banking?
            Volkswirtschaftliche Reihe

2009/3   EICHFELDER, Sebastian
            Tax compliance costs
            Volkswirtschaftliche Reihe

2009/4   SALIM, Claudia
            Optional linear input prices in vertical relations
            Volkswirtschaftliche Reihe

2009/5   BUSCH, Ulrike / Dieter NAUTZ
            Controllability and Persistence of Money Market Rates along the Yield Curve
            Volkswirtschaftliche Reihe

2009/6   BÖNKE, Timm / Carsten SCHRÖDER
            The German spatial poverty divide
            Volkswirtschaftliche Reihe

2009/7   BESTER, Helmut
            Investments and the Holdup Problem in a Matching Market
            Economics

2009/8   MELLER, Barbara / Dieter NAUTZ
            The Impact of the European Monetary Union on Inflation Persistence in the Euro Area
            Economics

2009/9   KREMER, Stephanie / Alexander BICK / Dieter NAUTZ
            Inflation and Growth
            Economics

2009/10   SCHÖB, Ronnie
            Climate Policy
            Economics

2009/11   KEREKES, Monika
            Growth Miracles and Failures in a Markov Switching Classification Model of Growth
            Economics