

Kerekes, Monika

**Conference Paper**

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

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Convergence and Structural Change, No. D5-V1

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Kerekes, Monika (2010) : Growth Miracles and Failures in a Markov Switching Classification Model of Growth, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010: Ökonomie der Familie - Session: Convergence and Structural Change, No. D5-V1, Verein für Socialpolitik, Frankfurt a. M.

This Version is available at:

<https://hdl.handle.net/10419/37334>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

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

Monika Kerekes\*<sup>†</sup>  
Freie Universität Berlin

February 2010

## Abstract

This paper interprets economic growth as a sequence of transitions between four growth regimes: crisis, stagnation, stable growth, and miracle growth. Countries experience distinct patterns of growth because they switch between the growth regimes with different frequencies. The aim of this paper is to identify countries with similar growth dynamics in order to determine the driving forces behind different within-country growth experiences. To that end, a traditional Markov switching model is estimated and extended by a classification mechanism that endogenously assigns countries that feature similar dynamics into the same, and countries that feature distinct dynamics into different clusters. Three clusters of countries are obtained: successful countries that are characterized by persistent periods of stable or miracle growth, moderately successful countries where growth is frequently interrupted by periods of stagnation and failing countries that stand out for their highly volatile growth rates with frequent periods of stagnation and crisis. It turns out that both a well-educated workforce and a geographically advantageous location reduce the probability of countries to belong to the cluster of failing countries, even if institutions are not well developed. However, in order to belong to the cluster of successful countries well-functioning institutions are essential.

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

---

\*Contact address: Freie Universität Berlin, Institute for Eastern European Studies, Garystr. 55, 14195 Berlin, Germany. Phone +49-30-838-52085, Fax +49-30-838-52072, e-mail: kerekes@wiwiss.fu-berlin.de.

<sup>†</sup>I would like to thank Giacomo Corneo and Wolfram Schrettl for helpful comments. All remaining errors are mine.

# 1 Introduction

In recent years it has been increasingly recognized that most countries do not grow at a steady rate, but rather experience substantial shifts in growth over time. Therefore, in order to gain a more thorough understanding of the growth process of countries, it is necessary to shift the focus of growth analysis away from explaining differences in the cross-country average growth rates to explaining the growth dynamics within countries, i. e. the transitions from phases of low growth to phases of high growth and vice versa (Pritchett, 2003, 2000; Jones and Olken, 2008; Hausmann et al., 2005). This paper extends the regime switching growth framework proposed by Pritchett (2003) and Jerzmanowski (2006) by a clustering mechanism that allows a more detailed analysis of the forces driving the within country growth dynamics.

The basic idea of the regime switching growth framework is to map different phases of economic growth in growth regimes or growth states.<sup>1</sup> For instance, a country experiencing very high growth rates for a decade (growth miracle) is in a different growth regime than a country stagnating for a decade. Hence, a growth regime is a state of nature in which a certain growth behavior prevails. In a world with several growth regimes countries can switch from one regime to another. For instance, bad political decisions might propel a former growth miracle into deep stagnation and thus induce a growth regime change. In this world, the current growth rate of a country is determined by the growth regime in effect and the economic forces within the state. The average long-run growth rate, however, depends both on the available growth regimes and the transitions between them. Suppose that a country can switch between regimes of fast growth, stagnation and crises. The long-run growth rate of this country is crucially determined by how it switches between the regimes and in particular by how much time it spends in each of them.

In the suggested framework countries show different patterns of growth depending on the way they switch between growth regimes. Pritchett (2000) shows that a multitude of growth patterns is observable, which implies that countries switch differently between growth regimes. These differences can be accounted for by introducing transition probabilities (i. e. probabilities to switch from one regime to another) and by letting them differ across countries. While every country can in principle switch to every available growth regime, the transition probabilities determine how frequent and how lasting a switch to a certain regime is. Therefore, in this framework it is ultimately the transition probabilities that are responsible for the different growth pattern across countries. A natural question then concerns the determinants of the transition probabilities. Obviously, the transition probabilities have to depend on long-run differences ("growth fundamentals") between countries rather than short-term events (e. g. an unfavorable change in the terms of trade) because the latter would be reflected by the actual switch of the growth regime. However, whether the transition probabilities depend on institutions, geography or human capital to name just a few conceivable growth fundamentals is an open question. Therefore, instead of conditioning the transition probabilities on a preselected growth fundamental, the regime switching framework is enriched by a clustering framework,<sup>2</sup> whereby those countries whose growth patterns can be described by similar transition probabilities are grouped together. The resulting clusters can then be compared along different dimensions in order to learn more about the driving forces behind different dynamics of growth across countries.

In order to identify the growth regimes, the transition between them and the clusters of countries with similar transition dynamics, a Markov switching classification model is estimated using the growth rate of GDP per capita in purchasing power parity for 84 countries over the period 1962 – 2002. Four growth regimes are identified: a stable growth regime, which is characterized by steady growth around two percent, 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 highly volatile growth behavior ranging from double-digit negative to double-digit positive growth rates. The countries are endogenously divided into three clusters:

---

<sup>1</sup> The terms growth regimes and growth states are used interchangeably in this paper.

<sup>2</sup> The terms classification and clustering are used interchangeably in this paper.

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 alternate between periods of stagnation and steady growth. Countries in the third cluster spend most of their time either in stagnation or in crisis and may therefore be referred to as growth failures. Based on these clusters a descriptive and quantitative analysis attempts to disentangle those growth fundamentals that can best explain the differences in the dynamics of growth. In addition to being open and well-educated, a distinctive feature of successful countries is the prevalence of trust-inspiring institutions. Neither moderately successful nor failing countries are characterized by a similar quality of institutions. However, compared to countries in the third cluster moderately successful countries have either a better educated workforce or simply benefit from a favorable geographic location. Such positive features appear to offset some of the negative implications associated with unreliable institutions.

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 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), Alfo et al. (2008) taking the "classical" estimation approach and in Ardic (2006) taking the Bayesian estimation approach. A related latent class model has been estimated by Bos et al. (2010). Other clustering methods such as projection pursuit (Desdoigts, 1999; Kourtellis, 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 that has previously mainly been used for automatic speech recognition 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 a potentially endogenous measure for the quality of institutions (Glaeser et al., 2004).<sup>3</sup> 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) and Rodrik (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. It therefore by construction handles both problems

---

<sup>3</sup> 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).

simultaneously. The results offer important hints at control variables that should be included when studying growth accelerations.<sup>4</sup>

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

## 2 Framework of Analysis

### 2.1 An encompassing growth model

Traditional growth theories are not able to explain the whole range of growth phenomena that are observed in reality. According to Pritchett (2003) an encompassing growth model should be able to explain the divergence in absolute and relative income levels that occurred in the last 150 years, the sustained steady growth rates observed in industrial countries, the extremely rapid growth rates of newly industrializing countries and the zero growth experiences of many poor countries. Moreover, a comprehensive model of growth should account for the observed instability of growth rates in developing countries as well as for the empirically established instability of estimated parameters. At present, there is no unified growth theory that can address all of these issues simultaneously. In order to overcome the limitations of each single growth theory, Pritchett (2003) suggests to define different growth regimes and assume that within each regime a well defined growth model governs the growth process. Different patterns of growth in this framework arise not only from transition dynamics within a state, but also from countries' switching 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.

Pritchett (2003) suggests a regime switching growth model comprising six growth regimes: The first regime 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,<sup>5</sup> 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 industrialized 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. The transitions between the growth regimes are captured by attaching transition probabilities to each state. Pritchett argues that these may depend on initial conditions, history, policies and institutions. Moreover, the transition probabilities are also determined by the sequence of previous growth regimes. 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).

### 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.

---

<sup>4</sup> A further study concerned with the sustainability of growth accelerations is Berg et al. (2008).

<sup>5</sup> 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).

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.<sup>6</sup> Formally, the growth rate in each state is given by

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

where  $y_{tk}$  is the growth rate of country  $k$  in period  $t$  and where  $s_t$  indicates the growth regime in effect. The growth rate in each period is influenced by a regime-specific random shock  $\varepsilon$ .<sup>7</sup> The evolution of the state variable is governed by the following first-order Markov chain, where

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

denotes the probability of changing from state  $i$  in the previous period to state  $j$  in the current period.<sup>8</sup> The first-order Markov chain assumption ensures that only the state in the last period, 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.

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

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

Jerzmanowski (2006) estimates the Markov switching model over the period 1962 – 1994 for 89 countries and obtains four growth states (see table 1): state one reflects stable growth with steady state growth rates around 2 %, state two is a stagnation state with zero growth in the steady state, state three is a crises state characterized by a large volatility of the 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 the new state two. Both changes can be attributed to the fact that income levels are no longer a defining element of the states. Regarding the transition probabilities, Jerzmanowski finds that better institutions lead to a greater persistence of the stable and the miracle growth regimes whereas poor institutions increase the persistence of stagnation and crisis. The main difference between fast and slowly growing countries is not the inability of slow growers to initiate growth, but the inability to sustain it.

---

<sup>6</sup> A more comprehensive definition is given in section 5.

<sup>7</sup> 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 et al. (2008) 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 defensible 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.

<sup>8</sup>  $\sum_j p_{ij} = 1$  holds. Cf. section 3.

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.

The results of the estimation are very appealing.<sup>9</sup> Apart from the good interpretability of the states, Jerzmanowski (2006) shows that the estimated regime changes often coincide with important political or economical events. Countries spend reasonable amounts of time in each of the regimes. One shortcoming of the estimation procedure, however, concerns the decision to condition the transition probability on the quality of institutions only. 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. In this case, the estimates would be inconsistent (Kim et al., 2008).<sup>10</sup> The usual remedies in the presence of endogenous variables such as the use of instrumental variables or control functions are not feasible for the current problem because both procedures would lead to even higher levels of non-linearity. In contrast to that, the endogenous clustering approach suggested in the following avoids the potential endogeneity and omitted variable problem without increasing the non-linearity of the estimation. Moreover, the influence of a whole set of variables on the transition probabilities can be analyzed.

The model design in this paper resembles that by Jerzmanowski (2006). The growth rates in each growth regime are modelled as AR(1) processes, i. e.

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

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 country-specific. Rather, countries are grouped into  $m$  clusters, whereby each cluster is characterized by distinct transition probabilities. Let  $C_m$  denote membership in cluster  $m$ . The transition probabilities conditional on belonging to cluster  $C_m$  are defined as

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

<sup>9</sup> One might object that business cycle fluctuations instead of long-run regime changes are identified due to the use of annual data. However, since a cross-section of countries is used, the different growth performances across countries are likely to dominate the differences between expansion and recession. Moreover, the analysis in any case provides a link between short(er)-run fluctuations and long-run economic performance (Jerzmanowski, 2006, p. 367). These conjectures are also supported by the obtained smoothed state probabilities in both Jerzmanowski's model and the one estimated in this paper. For instance, most industrialized countries of 1960 do not report growth regime changes even though these countries have experienced business cycle fluctuations.

<sup>10</sup> 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.

Cluster membership is endogenously determined in the estimation process. Compared to Jerzmanowski (2006) this model specification is less restrictive in the sense that countries with the same quality of institutions may feature different transition probabilities. The specification is more restrictive in the sense 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 fundamental forces driving the transition dynamics in the regime switching growth framework.

### 3 Methodology

The model proposed in the previous section 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, first the estimation strategy for simple Markov switching models will be summarized briefly. After that the idea of the clustering mechanism is introduced and the formulas for the parameter estimation are derived. Finally, some methodological issues are discussed.

#### 3.1 The Basic Markov Switching Model

The basic  $Q$ -state Markov switching model is given by

$$y_t = x_t' \beta_{s_t} + \varepsilon_t^{s_t}, \quad t = 1, \dots, T. \quad (5)$$

The evolution of the dependent variable  $y$  depends on the observed exogenous variables  $x$ ,<sup>11</sup> which may include autoregressive terms, *and* the contemporaneous state of the model  $s_t \in \{1, \dots, Q\}$  because the marginal effects of the exogenous variables equalling  $\beta_{s_t}$  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  $\mathbf{P}$

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{Q1} \\ p_{12} & p_{22} & \ddots & p_{Q2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1Q} & p_{2Q} & \cdots & p_{QQ} \end{bmatrix}, \quad (6)$$

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 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 is given by<sup>12</sup>

<sup>11</sup>  $x_t$  may be a  $(v \times 1)$  vector of exogenous variables,  $v \geq 1$ . Accordingly,  $\beta_{s_t}$  may be a parameter or a vector of parameters.

<sup>12</sup> A more detailed derivation of the complete-data log likelihood function is delegated to appendix A.



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

$f(\cdot)$  denotes the conditional density of  $y_t$ .  $\rho_j$  is a shortcut for  $P(s_1 = j | \psi_0, \theta)$ .  $\psi_t$  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  $\theta$ . Hence,  $\theta$  consists of  $(\beta_1, \dots, \beta_Q)$ ,  $(\sigma_1^2, \dots, \sigma_Q^2)$  and the transition probabilities. The ergodic distribution of  $\mathbf{P}$  is used to predict the initial probabilities  $\rho_j$ ,  $\forall j = 1, \dots, Q$ . 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).<sup>13</sup> 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  $\theta$  conditional on this best guess, the so called maximization step. Any best guess of  $s_t$  necessarily depends on  $\theta$ . Therefore, the estimation and maximization steps are iterated until convergence of  $\theta$  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  $\theta$  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 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  $t^{\text{th}}$  observation has been generated by regime  $i$  conditional on a parameter value  $\theta$  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  $\eta_t$  and that the  $Q$  calculated 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

---

<sup>13</sup> 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).

iterating forwards on the following two equations:<sup>14</sup>

$$\xi_{t|t} = \frac{\xi_{t|t-1} \odot \eta_t}{1'(\xi_{t|t-1} \odot \eta_t)} \quad (8)$$

$$\xi_{t+1|t} = \mathbf{P} \cdot \xi_{t|t}. \quad (9)$$

In order to start the filter,  $\xi_{1|0}$  is assumed to equal the ergodic distribution implied by the transition probabilities.<sup>15</sup>

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  $Q$  calculated smoothed probabilities of period  $t$  are collected in the  $(Q \times 1)$  matrix  $\xi_{t|T}$ .

$$\xi_{t|T} = \xi_{t|t} \odot \{\mathbf{P}' \cdot [\xi_{t+1|T} \odot \xi_{t+1|t}]\} \quad (10)$$

The backwards recursion is started by using  $\xi_{T|T}$  obtained from the filtered state probabilities.<sup>16</sup>

### 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 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.<sup>17</sup>

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 first the complete-data log-likelihood function for this revised problem is established 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  $m = 1, \dots, M$  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  $\theta_m$  and summarize all  $\theta_m$  in the parameter vector  $\theta$ . 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  $\theta$  being replaced by the appropriate  $\theta_m$ . 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.  $C_m(\mathcal{O}^k)$  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

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

<sup>15</sup> Cf. Hamilton (1994, p. 693) for further implementation possibilities.

<sup>16</sup> 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).

<sup>17</sup> Panel Markov switching models are not very common in the economics literature. Some examples are Asea and Blomberg (1998) and Chen (2007b,a). 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 7.

information is given by<sup>18</sup>

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

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  $\theta$  the smoothed state probabilities for each observation in each cluster are calculated (expectation step). 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.<sup>19</sup> 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  $\theta$  is met. Usually, cluster memberships only change in the first iterations and remain constant afterwards.<sup>20</sup>

### 3.3 Parameter Estimation

In this section the required formulas for the estimation of the model are derived. The estimation is carried out by implementing the algorithm in Matlab. For convenience the growth model from section 2 is briefly repeated. There are  $Q$  different growth regimes that countries can visit during their growth process. Within each regime the growth rate evolves according to an AR(1) process with state-specific coefficients. The dynamics of growth between countries differ because they visit the growth regimes with different frequencies. This feature is captured in the accompanying transition matrix. Each country belong to one of  $M$  clusters, which feature distinct transition matrices. The regime-specific coefficients and variance, on the other hand, are equal across regimes. Formally the growth rate in each regime is given by:

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

Each of the  $M$  clusters is characterized by a unique transition probability matrix  $\mathbf{P}^m$ :

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

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

---

<sup>18</sup> Owing to the panel context it now holds that  $\mathcal{Y}_T = \{\mathcal{Y}_{1T}, \dots, \mathcal{Y}_{KT}\}$ , 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.

<sup>19</sup> Cf. Appendix B for a more detailed explanation.

<sup>20</sup> 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

$$\begin{aligned}
\mathcal{L}(\mathcal{Y}_t, \mathcal{S}_T, \mathcal{C} | \mathcal{X}_t, \theta) &= \sum_{k=1}^K \left\{ \sum_{m=1}^M \mathcal{C}_m(\mathcal{O}^k) \left\{ \sum_{j=1}^Q I(s_t = j) \right. \right. \\
&\quad \left. \left( -\frac{1}{2} \log(2\pi) - \frac{1}{2} \sigma_j^2 - \frac{1}{2} \frac{(y_{1k} - \alpha_j - \beta_j y_{0k})^2}{\sigma_j^2} + \log \rho_j \right) \right. \\
&\quad \left. + \sum_{t=2}^T \left[ \sum_{j=1}^Q I(s_t = j) \left( -\frac{1}{2} \log(2\pi) - \frac{1}{2} \sigma_j^2 - \frac{1}{2} \frac{(y_{tk} - \alpha_j - \beta_j y_{t-1k})^2}{\sigma_j^2} \right) \right. \right. \\
&\quad \left. \left. + \sum_{i=1}^Q \sum_{j=1}^Q I(s_t = j, s_{t-1} = i) \log p_{ij}^m \right] \right\} \left. \right\}. \tag{14}
\end{aligned}$$

In the following the EM algorithm for one iteration step 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  $\theta$  and a classification of the time-series into clusters is available. For the sake of clarity, neither estimated quantities nor iteration rounds are indicated separately. However, the accompanying notes clarify these issues. Once an updated estimate is available it continues to be used until the end of the iteration step. 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  $\theta$ , 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  $\theta$  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 cluster  $m$ . 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  $\rho_j^m$  corresponds to the probability of state  $j$  in the ergodic distribution associated with  $\mathbf{P}^m$ .

4. Summarize the coefficients  $\alpha_j$  and  $\beta_j$  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  $\kappa_j$  is obtained as

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

where

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

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

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

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

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, 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:<sup>21</sup> 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.<sup>22</sup> 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, both Pritchett (2003) and 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

---

<sup>21</sup> 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 Titterton, 1998).

<sup>22</sup> Several methods are shown in Biernacki et al. (2003), and McLachlan and Krishnan (1997).

sampling variation (Hansen, 1992; Garcia, 1998).<sup>23</sup> 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 is 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.<sup>24</sup> However, since each time-series is deterministically assigned 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 average accuracy of the model in the present context can be analyzed via simulation. To that end, the parameter estimates for each number of clusters have been used to simulate time-series<sup>25</sup> 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.<sup>26</sup>

As expected, the classification works better, the more observations are available. The classifica-

---

<sup>23</sup> 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.

<sup>24</sup> 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)).

<sup>25</sup> Slight adjustments of the transition probability matrix have been made if a rounded entry equalled 0.000

<sup>26</sup> 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.

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.

tion 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.

## 4 Empirical Results

### 4.1 Parameter Estimates

In this section the model is taken to the data. The data set consists of the annual growth rates of real GDP per capita measured in 2000 international purchasing power parity dollars 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.<sup>27</sup> 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  $M = 1, \dots, 4$  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.

<sup>27</sup> 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
<i>BIC</i>	5.6370	5.5777	5.5690	5.5710
Modified <i>BIC</i>	5.6382	5.5803	5.5729	5.5761
<i>CAIC</i>	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.

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 episodes of extremely positive growth rates. Therefore, the most outstanding feature of 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 the name "crisis regime" is retained for ease of comparison. The second state corresponds to Jerzmanowski's stagnation state. This regime features a constant that is not significantly different from zero and significant but rather small positive autocorrelation. The standard error of this regime is much smaller than that of the crisis 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, positive 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 experience 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



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.

developed countries. It is 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 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. This cluster is referred to 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 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 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

Table 5: Transition Probability Estimates and Country Classification

Successful Growth Cluster					Moderately Successful Growth Cluster					Growth Failures				
State	State	State	State	State	State	State	State	State	State	State	State	State	State	State
1	2	3	4	4	1	2	3	4	4	1	2	3	4	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					Ergodic					Ergodic				
Distribution	0.0018	0.0587	0.7682	0.1713	Distribution	0.0169	0.3667	0.4525	0.1638	Distribution	0.1759	0.5685	0.1939	0.0616
<u>Africa</u>					<u>Africa</u>					<u>Africa</u>				
South Africa					Burkina Faso					Algeria				
					Egypt					Benin				
<u>Asia</u>					Guinea					Burundi				
Hong Kong					Kenya					Cameroon				
Japan					Morocco					Chad				
South Korea					Tunisia					Cote d'Ivoire				
United States										Ethiopia				
<u>North America</u>										Ghana				
Canada										Madagascar				
Guatemala										Malawi				
El Salvador										Mali				
Brazil										Mozambique				
Colombia										Niger				
Paraguay										Nigeria				
Oceania										Rwanda				
Australia										Senegal				
New Zealand										Tanzania				
										Togo				
<u>Europe</u>										Uganda				
Austria										Zambia				
Belgium										Zimbabwe				
France														
Greece														
Ireland														
Italy														
Netherlands														
Norway														
Portugal														
Spain														
Sweden														
Switzerland														
United Kingdom														

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

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. 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.

This section concludes with a short reflection 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. The present approach is purely driven by the observed growth processes so that 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 6 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 12 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 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 has achieved the growth rate despite frequent periods of stagnation thanks to relatively fre-

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.

quent periods of miracle 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 through 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).<sup>28</sup>

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

<sup>28</sup> A table containing the average regime probabilities for all countries can be found in appendix C.

Table 7: Counterfactual Smoothed State Probabilities

Country	Cluster	Crisis	Stagnation	Stable Growth	Miracle Growth
China	<b>1</b>	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	<b>1</b>	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
	<b>3</b>	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
	<b>3</b>	0.15	0.48	0.31	0.06
Malaysia	<b>1</b>	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	<b>1</b>	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
	<b>2</b>	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.

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 vein, 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 a combination of stagnation and 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.

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 led to a year of political turmoil. 1997 marks the Asian Financial Crisis that had severe short-run impacts on

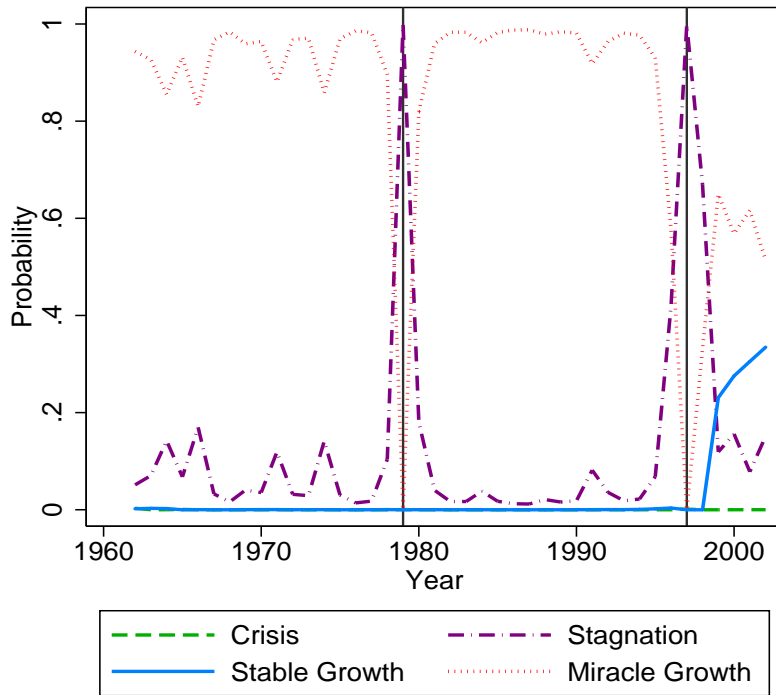


Figure 1: Average Regime Probabilities: Korea

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.

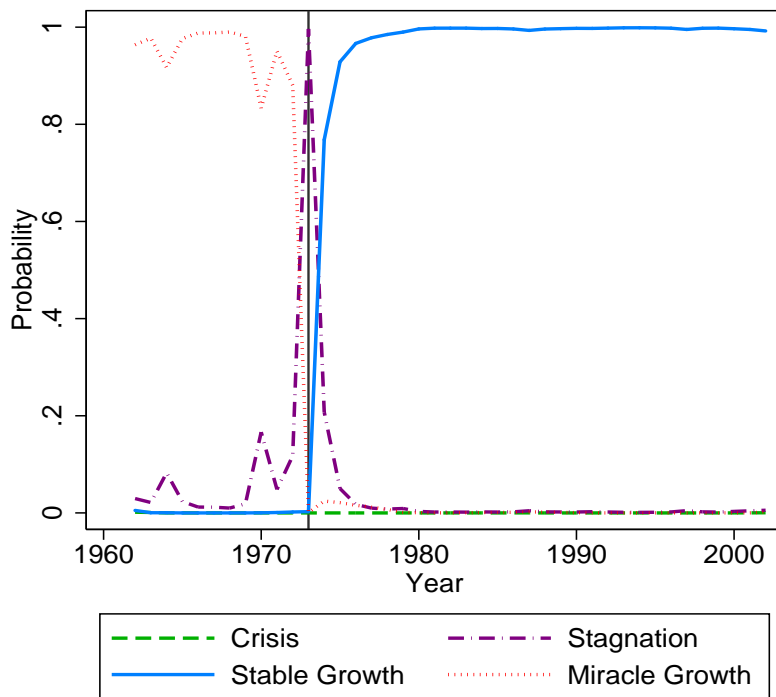


Figure 2: Average Regime Probabilities: Japan

Unfortunately, the 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 reforms 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.

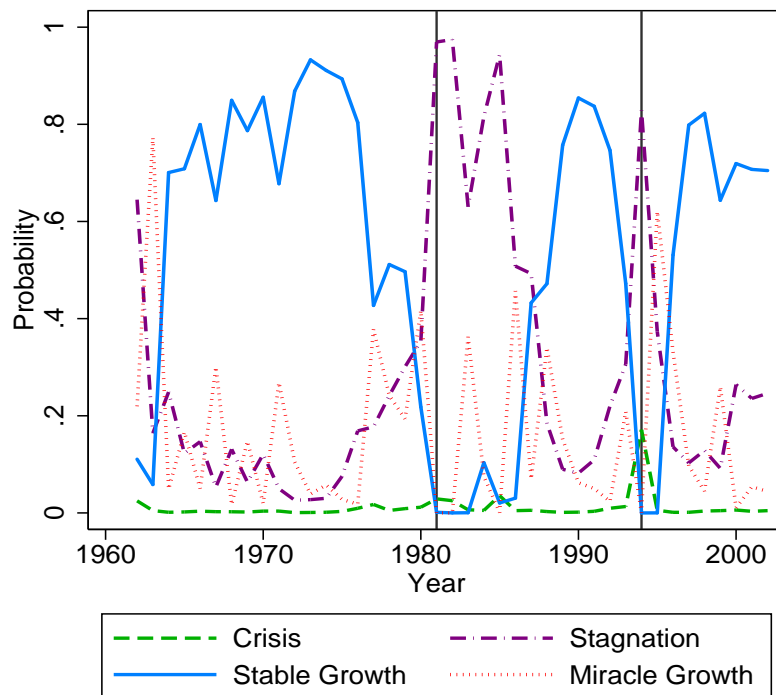


Figure 3: Average Regime Probabilities: Mexico

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 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 result 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.

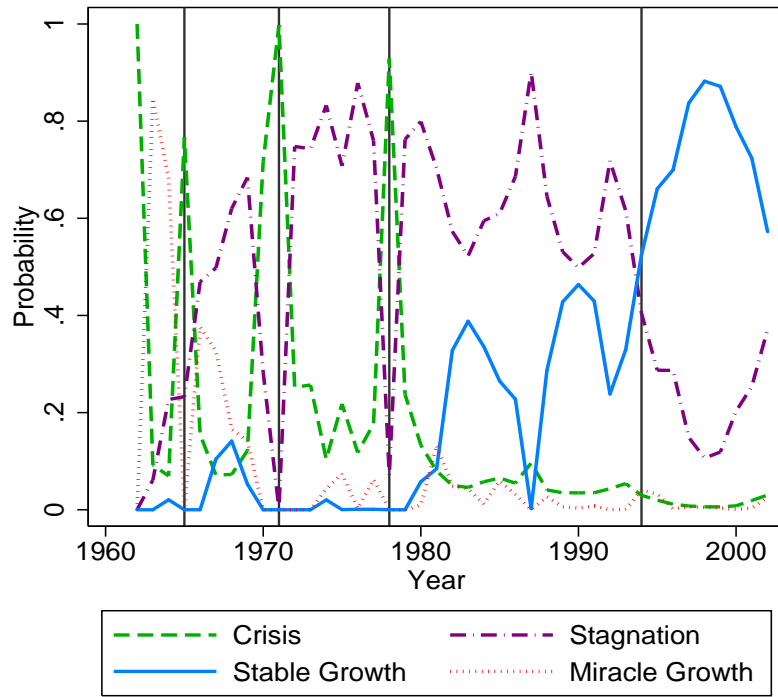


Figure 4: Average Regime Probabilities: Algeria



## 5 Accounting for the Assignment to Clusters

This section asks whether the identified clusters of countries are associated with differences in underlying growth fundamentals. In his analysis Jerzmanowski (2006) had to postulate a relationship between growth fundamentals and dynamics that depended on one variable in order to deal with the high non-linearity of the estimation problem. Contrary to that, the clustering technique in this paper allows comparisons of clusters along a multitude of dimensions in a second step, where the non-linearity of the estimation is no longer an issue.<sup>29</sup> Following Berthelemy (2006), the identified clusters are compared along indicators for three fundamental issues: the ability of the economies to accumulate reproducible factors, basic structural features and institutions. The main focus is on differences in the beginning of the sample period so as to mitigate problems related to reverse causality.<sup>30</sup>

### 5.1 A Descriptive Approach

Do country clusters differ with regard to the accumulation of production factors and if so why? The average saving and investment rates in each cluster of countries provide a first indication whether such differences exist. Different saving and investment propensities may result from differences in demographics. For instance, a high growth rate of the population deters savings if people are living close to subsistence levels (Ben-David, 1998). Moreover, the capital stock is diluted across a large number of persons and may thus further impede additional savings by keeping the capital intensity of the economy low (Nelson, 1956). A minimal level of human capital may be required in order to make investments into physical and human capital worthwhile (Azariadis and Drazen, 1990).

Table 8 reports the average investment and saving rates, the population growth rates, the dependency ratios (the ratio of the number of people younger than 15 or elder than 64 to the number of people between 15 and 64 of age), average years of schooling and the percentage of the population having completed at least primary school for each of the clusters in 1960 or 1965 depending on the availability of data. For each pair of clusters, a group mean comparison test is conducted assuming unequal variances and using Satterthwaite's degree of freedom adjustment. A significant difference at the five percent level is indicated by a superscript.<sup>31</sup> The dynamically growing countries in cluster one exhibit significantly higher saving and investment rates and significantly lower population growth rates and dependency ratios than their counterparts, corroborating the idea that successful development and growth are fundamentally linked to the ability of economies to accumulate physical capital. However, the differences between cluster two and cluster three are not significant, indicating that other factors, too, must be important. One of those factors appears to be the availability of human capital. Both the average years of schooling and the percentage of the population having completed at least primary school is significantly lower in the growth failure than in the other two clusters. However, moderately successful countries also exhibit significantly less education than the countries in cluster one. Therefore, sluggish development seems to be related to too little education.

Next, the influence of favorable structural features on the assignment to clusters is examined. A strong financial system potentially facilitates the capital accumulation process because it pools funds and allocates them towards investments offering the highest returns (King and Levine, 1993b,a). A certain degree of openness to trade may be favorable because it allows countries to pursue export-led growth strategies. Geographic conditions also affect the structure of economies.

---

<sup>29</sup> There are no adjustments to account for the uncertainty inherent in the clustering results. Similar approaches have been used by Durlauf and Johnson (1995) and Ardic (2006).

<sup>30</sup> However, since countries are assumed to belong to one cluster once and for all (in particular also before 1961), one might even argue that the initial conditions are endogenous. In this case, the following analysis should not be interpreted causally at all.

<sup>31</sup> For reasons of clarity the significance of the differences for each cluster are only tested against clusters with a larger number. It should be noted that due to the repeated testing procedure the type I error rates are inflated. For instance, in order to find an overall type I error rate less or equal to five percent, each difference in the means would have to be significant at the  $0.05 / 6 = 0.0083$  significance level (Shaffer, 1995).

Table 8: Factor Accumulation

	Investment Rate 1960	Saving Rate 1965	Population Growth Rate 1960	Dependency Ratio 1960	Average Years of Schooling 1960	Primary School Com- pleted 1960
Cluster 1	19.78 <sup>2</sup> (1.59) 31	22.71 <sup>2,3</sup> (2.02) 21	1.95 <sup>3</sup> (0.25) 31	0.71 <sup>2,3</sup> (0.03) 30	5.50 <sup>2,3</sup> (0.47) 31	54.74 <sup>2,3</sup> (4.24) 31
Cluster 2	13.33 (1.75) 24	14.96 (1.88) 19	2.34 (0.17) 25	0.83 (0.03) 25	2.88 <sup>3</sup> (0.46) 25	29.93 <sup>3</sup> (4.44) 23
Cluster 3	12.71 (3.14) 28	14.63 (2.47) 23	2.53 (0.12) 28	0.89 (0.02) 28	1.43 (0.24) 28	13.73 (1.68) 25

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 Sattersthaite's degree of freedom adjustment. The investment and population growth rates are taken from the Penn World Tables 6.2 (Heston et al., 2006), the saving rates and the dependency ratios are based on the World Development Indicators 2007 (WorldBank, 2007). The average years of schooling in 1960 refer to the non-studying population aged 15-64 and are taken from Cohen and Soto (2001). The percentage of the population having completed at least primary school is taken from Barro and Lee (1993, 2001). Linear regression methods and the additional data set provided by Lutz et al. (2007) are used in order to extend the data coverage of the latter two datasets as much as possible.

Table 9: Structural Features

	Private Sector Credit to GDP 1965	Openness Dummy 1970	Years of Liberal- ized Trade 1960 – 2002	Landlocked Country	Percentage of Land Within 100km to the Coast	Percentage of Land in Tropical Climate
Cluster 1	36.40 <sup>2,3</sup> (4.82) 28	0.74 <sup>2,3</sup> (0.08) 31	33.26 <sup>2,3</sup> (2.34) 31	0.10 <sup>3</sup> (0.05) 31	0.52 <sup>3</sup> (0.07) 31	0.30 <sup>2,3</sup> (0.08) 31
Cluster 2	19.04 <sup>3</sup> (2.50) 22	0.20 (0.08) 25	17.20 <sup>3</sup> (2.30) 25	0.12 <sup>3</sup> (0.07) 25	0.52 <sup>3</sup> (0.08) 25	0.53 <sup>3</sup> (0.10) 25
Cluster 3	11.93 (1.34) 23	0.04 (0.04) 28	7.61 (1.46) 28	0.36 (0.09) 28	0.14 (0.03) 28	0.85 (0.06) 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 Sattersthaite's degree of freedom adjustment. Credit to GDP is taken from the World Development Indicators 2007 (WorldBank, 2007). The openness indicators are taken from Sachs and Warner (1995) and Wacziarg and Welch (2008), the geographic variables from Gallup et al. (1999).

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 amount of credit directed to the private sector as a percentage of GDP is used to capture the state of the financial system. Following Sachs and Warner (1995) and Wacziarg and Welch (2008) a country is classified as open (and the corresponding dummy

variable takes on the value one) if it does not impose too many tariff and non-tariff barriers and if the government is not intervening too much either in trade or in foreign exchange activities.<sup>32</sup> In addition, openness is measured as the average years a country has been liberalized in the period 1960 – 2002 based on Wacziarg and Welch (2008). A dummy variable indicating countries without access to the sea, the percentage of land within 100 km to the coast and the percentage of land in tropical climate are used to measure the geographic conditions of the countries.

Table 9 reveals that the successful countries had significantly better structural features at the beginning of the sample period than moderately successful countries and these again had better features than countries in the growth failure cluster. Regarding those factors that can be influenced politically, successful growth countries have been open for longer periods of time than moderately successful countries and had a more developed financial sector. With regard to geographic conditions, both clusters feature similar conditions, which are much more favorable than those for cluster three. One third of the least successful countries does not have access to the sea at all and less than one third of the populations lives within one hundred kilometers to the coast. However, 85 % of the land is tropical. Thus, while it is possible to overcome disadvantages associated with geographic conditions (as proven by the landlocked or tropical countries that find themselves in clusters one or two nevertheless), less favorable geographic conditions make growth harder to achieve.

Table 10: Quality of Institutions

	Government Antidiver- sion Index (GADP)	Voice and Account- ability (VA1996)	Independence of Judicity	BERI index 1972	BERI index 1995
Cluster 1	0.81 <sup>2,3</sup> (0.03) 30	0.71 <sup>2,3</sup> (0.14) 31	0.76 (0.05) 26	11.36 <sup>2,3</sup> (0.41) 21	10.56 <sup>2,3</sup> (0.38) 23
Cluster 2	0.57 <sup>3</sup> (0.03) 24	0.03 <sup>3</sup> (0.16) 25	0.82 (0.07) 15	7.27 (0.86) 12	8.33 (0.54) 10
Cluster 3	0.49 (0.01) 28	-0.54 (0.14) 28	0.73 (0.09) 15	7.22 (0.55) 5	6.84 (0.58) 5

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 Sattersthaite's degree of freedom adjustment. The data is taken from Hall and Jones (1999), Kaufmann et al. (2008), La Porta et al. (2004) and Knack and Keefer (1995).

In recent years, the idea that well designed institutions are the ultimate cause for growth has gained acceptance (Acemoglu et al., 2001, 2003; Collier, 2006). Since Jerzmanowski (2006), too, argues in this spirit it is particularly interesting to see whether the identified clusters indeed differ in this respect. Table 10 reports four indicators for the quality of institutions (GADP - government antidiversion index (Hall and Jones, 1999), VA - voice and accountability indicator (Kaufmann et al., 2008), independence of judicity (La Porta et al., 2004), BERI - Business Environmental Risk Intelligence (Knack and Keefer, 1995)<sup>33</sup>), all of which are defined such that a larger value indicates better quality.<sup>34</sup> Most institutional quality indicators have only been gathered from 1990 onwards,

<sup>32</sup> The openness indicator is reported for 1970 because this way it includes those countries that liberalized in the Kennedy GATT round from 1964 – 1967. However, a sizeable number of countries only liberalized in the Tokyo round from 1973 – 1979. Therefore, the average years a country has been open is also reported, even though this variable should not be used for causal interpretation because trade liberalization decisions almost surely depend to some extent on economic development.

<sup>33</sup> Thank you to Knack and Keefer (1995), who kindly provided their data collection.

<sup>34</sup> A more detailed description is delegated to appendix D.

so that they refer to the state at the end of the sample raising the issue of causality. The only exception is the BERI indicator, which has been published in 1972 for the first time. All indicators with the exception of judicial independence show that the average institutional quality in successful countries significantly exceeds that of the other two clusters. The indicators measured at the end of the sample period also suggest that institutional quality in moderately successful countries is better than in cluster three countries. However, the differences are smaller than the differences between cluster one and cluster two countries and the BERI index does not support the conclusion. The latter may be a consequence of the small number of observations for countries belonging to cluster three, though.

Summing up, countries belonging to the successful growth cluster show better growth fundamentals in every category analyzed. This result is not particularly surprising because cluster one consists to a non-negligible extent of industrialized countries and to some extent the growth fundamentals identified in the literature simply reflect the differences between developed and developing countries. However, precisely because countries in cluster one have done well in all the categories, it is difficult to discern the driving force behind the assignment to clusters. In this respect, the differences between clusters two and three are potentially more useful, since these clusters differ in some respects, but not in all. The most likely candidates that distinguish moderately successful from failing countries are differences in human capital and education, geographic differences, different trade liberalization decisions and differences in the quality of institutions. As already mentioned in section 2, one potential shortcoming of the suggested clustering method is that it focuses solely on growth rates without taking account of differences in income levels. Pritchett (2003), however, argues that the growth process of industrialized countries differs from that of developing countries. For this reason, the descriptive statistics of this section are reproduced in Appendix E for the case that the developed countries of 1960 are placed into a separate cluster four. The main results are robust to this alteration: successful developing countries differ from less successful countries mainly due to advantages in education, openness and geography, and due to better institutions. Thus, the same variables are defined as the most likely candidates to explain the assignment to clusters.

## 5.2 Regression Results

To synthesize the foregoing discussion, a multinomial logit model is estimated using cluster three (growth failures) as the reference group. Since the number of observations in each cluster and overall is small, the results of the previous section are used to fit a parsimonious model. To start with, only the average years of schooling, the GADP indicator for the quality of institutions and the openness dummy for 1970 are included as explanatory variables, because these appear to be the most important variables that can be influenced by policy.<sup>35</sup> Geographic conditions are initially excluded in the hope that they do not influence the assignment to clusters over and above the policy variables. The results of the basic specification are found in column one of Table 11. For each explanatory variable the coefficient estimate, standard error and average marginal effect is reported.<sup>36</sup> The most important difference between moderately successful and cluster three countries seems to be the higher level of education in the former, whereas successful countries stand out due to their high quality of institutions. In this specification, openness is not significant in either of the sub-regressions and a Wald-test on the overall insignificance of this variable cannot be rejected. Dropping the openness dummy (column two) leaves the pseudo- $R^2$  essentially unchanged, but renders education significant in cluster one, too. The reason seems to be a high degree of collinearity between the three variables in cluster one.<sup>37</sup> The predictive capacity of both the basic and the reduced specification is reasonable for cluster one and cluster three, but disappointing with regard to cluster two.

---

<sup>35</sup> The openness dummy is preferred to the total years of liberalized trade because it is less likely to be influenced by the assignment to a certain cluster.

<sup>36</sup> Cf. Appendix F for a detailed discussion of the estimation model.

<sup>37</sup> Appendix G reports the correlation matrices for the regressors.

Table 11: Multinomial Logit Specifications

	(1)	(2)	(3)	(4)	(5)	(6)
Cluster 1						
Years of schooling 1960	0.426 (0.272)	0.480* (0.266)	0.241 (0.284)	0.789** (0.269)	0.674** (0.271)	0.576* (0.296)
Openness Dummy 1970	0.012 (1.276)	0.017	0.011	0.075	0.071	0.051 (1.285)
GADP	2.004 (1.276)	0.215	0.829** (4.163)	11.218** (3.892)	11.858** (4.410)	0.362
Coastal Land			1.104** (1.568)		3.435** (1.557)	.927* (1.635)
VA 1996			0.077	0.600 (0.627)	0.248 (0.670)	-0.052 (0.712)
				0.047	0.036	-0.006
Cluster 2						
Years of schooling 1960	0.479** (0.225)	.530** (0.223)	0.226 (0.257)	0.430* (0.240)	0.282 (0.256)	0.260 (0.269)
Openness Dummy 1970	0.055 (1.234)	0.060	0.017	0.010	-0.016	-0.005 (1.334)
GADP	-0.028 (3.413)	2.857 (3.221)	3.039 (3.896)	4.190 (3.896)		-0.193
Coastal Land	-0.176	-0.330	-0.298 (1.414)	4.150** (1.414)	4.405** (1.518)	4.292** (1.568)
VA 1996			0.391	0.432 (0.492)	0.446 (0.570)	0.470 (0.582)
VA 1996				0.028	-0.023	0.003
Pseudo $R^2$	0.319	0.302	0.370	0.240	0.308	0.351
$\chi^2$	57.332	54.237	66.477	43.007	55.277	62.965
p-value	0.000	0.000	0.000	0.000	0.000	0.000
Proportion correctly predicted						
Cluster 1	0.806	0.839	0.839	0.645	0.613	0.774
Cluster 2	0.400	0.320	0.560	0.120	0.360	0.480
Cluster 3	0.821	0.821	0.821	0.750	0.857	0.857

Standard errors are reported in parenthesis. \* denotes significance at the 10%-level, \*\* significance at the 5%-level. The third line for each variable reports the average partial effect. The number of observations in total is 82 (81 in (3) due to Tunisia). Cluster one contains 30, cluster two 24, cluster three 28 observations.

Using the reduced specification, additional explanatory variables referring to the capital accumulation process and the structural features of the economy (excluding geography) are introduced one by one. None of them is significant at the five percent level and the Wald test on the overall insignificance of the respective variables is always accepted. Regarding geography, neither the dummy for being a landlocked country nor the percentage of land in the tropics alters the results. However, the inclusion of the percentage of land within 100 km to the coast has a large impact. In both clusters, geography becomes significant whereas education loses its significance. The average marginal effect of education plummets in cluster two. While the changing contribution of education can again be attributed to multicollinearity problems<sup>38</sup>, the increased pseudo-R<sup>2</sup> and notably better predictive power for cluster two imply that the geographic conditions contain additional information previously not captured. Favorable geographic conditions appear to play an important role in shaping a positive environment for growth.<sup>39</sup>

So far, the quality of institutions as measured by the GADP indicator appears to exert a robust positive effect on the probability of a country to become a successful grower. This, however, is a special feature of the indicator and does not hold if the VA indicator is used instead (columns four to six). In this case, the predictive capacity of the model deteriorates considerably and the quality of institutions is never significant. Instead, openness and education become significant predictors for membership in cluster one. In cluster two the percentage of coastal land continues to drive the results.

In summary, the finding that significance is only ever attached to a combination of variables representing education, openness, the quality of institutions and favorable geographic conditions implies that these variables are at the core of differing growth dynamics between countries.<sup>40</sup> Yet, the high degree of collinearity between these variables prevents the identification of a single most important driving factor among these variables. The GADP indicator seems to be particularly useful to identify successful growers. Even though the index is measured relatively late and may therefore have been influenced by the dynamics of growth itself, it nevertheless seems to capture a much more important aspect of institutional quality than the voice and accountability indicator does, which is also measured at the end of the sample period. The latter indicator focuses on freedom, while the former emphasizes property rights and contractual reliability, two features that have been singled out as the most important institutions of economies (Acemoglu, 2009, chap.4.5). Therefore, successful development seems to require institutions that guarantee a large amount of certainty in conducting business. However, if institutions do not provide a certain threshold of security, the dynamics of growth can no longer be well predicted based on the quality of institutions alone. Rather, in this case it is important whether the country has other advantages such as a well educated work force or a favorable location that can to some extent compensate for the institutional disadvantages. In light of these results Jerzmanowski's (2006) choice to condition the transition probabilities of countries on the GADP index is probably the best possible choice given the restrictions due to the high nonlinearity of the problem. The approach reliably captures the differences in the transition probabilities between successful and less successful countries. However, since the GADP indicator has difficulties to differentiate between moderately successful and failing countries, the depicted differences in the transition matrices of less successful countries should not be taken literally. By conditioning on the quality of institutions alone, mitigating effects that a well educated workforce, an advantageous geographic location or openness can have in the presence of weak institutions are not accounted for and can lead to too gloomy or too kind a transition matrix estimate compared to reality.

---

<sup>38</sup> The correlation between education and land within 100 km to the coast is 0.57 in cluster two.

<sup>39</sup> The insignificance of additional explanatory variables also holds with this model specification.

<sup>40</sup> The results are qualitatively the same if the industrialized countries of 1960 are treated as a separate cluster. Cf. Appendix H.

## 6 Conclusion

The variation of growth rates within countries has been identified as an important research topic. This paper contributes to the discussion by interpreting the process of growth within the regime-switching framework proposed by Pritchett (2003) and Jerzmanowski (2006). In this framework economic growth is a process during which countries switch between the same distinct growth regimes. The frequency of transitions between the regimes is determined by a collection of transition probabilities. Empirically, different patterns of growth are observed, which implies that the transition probabilities differ across countries. In order to capture the most important differences in the transition probabilities without increasing the non-linearity of the associated estimation problem so much as to render the model intractable, the regime-switching framework is extended by a clustering mechanism. The model is estimated for 84 countries over the period 1962 – 2002. In the estimation process the growth regimes and transition probabilities are determined endogenously. Moreover, each country is endogeneously assigned to the cluster which maximizes the likelihood of observing the actual data.

The model identifies four distinct growth regimes and three clusters of countries. In the stable growth regime countries grow steadily with a long-run growth rate of around two percent. The miracle growth state features persistently fast growth around five to six percentage points per year and allows countries to embark on a catch-up process. The stagnation regime is characterized by zero growth on average and a low persistency of growth rates. Finally, the crisis state captures very large shocks, both positive and negative ones, although on average negative shocks are more frequent. The derived clusters of countries imply distinct patterns of growth for each of the groups. Countries in the first cluster are successful countries such as the industrialized and newly industrialized countries and some steadily growing developing countries. Their growth process is characterized by steady growth and prolonged growth spurts. The second and the third cluster consist mainly of developing countries. Countries in the second cluster are moderately successful growers that spend the majority of their time in the stagnation and the stable growth regime with approximately equal weight. Growth spurts in this cluster do occur, but they are short-lived compared to those of cluster one. Finally, countries in the third cluster are characterized by disappointing growth experiences. They usually find themselves either in stagnation or in crisis and exhibit a very large volatility of growth rates. The income gap between these countries and industrial leaders continues to diverge so that the countries may be called laggards or growth failures. Conforming to conventional wisdom, the majority of African countries is associated with this cluster.

Since ultimately the transition probabilities are responsible for the differences in the patterns of growth, it is instructive to analyze in what respects the resulting clusters of countries differ from each other. Such an analysis offers important insights into useful policy measures in order to improve the chances of sustainable growth. This paper investigates whether the clusters differ with respect to their ability to accumulate physical and human capital, to structural features of the economies and to the quality of institutions. While the successful countries of cluster one outclass the other two clusters in all of these aspects, the most outstanding feature is the prevalence of trust-inspiring institutions. In the absence of good institutions, the availability of human capital, favorable geographic conditions and openness can mitigate negative consequences of the resulting uncertainty and prevent countries from becoming a growth failure.

Based on the previous findings the following policy recommendations emerge. Developing countries should invest in education in order to enhance their human capital base, because an educated workforce seems to be good strategy to to mitigate negative effects that result from the often poorly designed institutions in developing countries. Since this measure is relatively easy to implement and does not generate too much resistance from vested interest groups, it is a useful first step in any development plan. Countries with favorable geographic conditions seem to have an inherent advantage in succeeding to reach a successful or moderately successful growth path. Most likely, this advantage is due to lower transportation costs that make it easier and more worthwhile for these countries to engage in trade or to attract foreign direct investment, and hence to benefit from an export-led growth strategy and the diffusion of technology. Therefore, countries that do not have these advantages should focus on the expansion of roads, and 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. Unfortunately, this measure is likely to be the most difficult to implement.

In a more technical sense, this paper confirms Jerzmanowski's (2006) choice to use the quality of institutions as the conditioning variable for estimating country-specific transition probabilities. This variable captures the distinctive dynamics between successful and less successful countries. However, the analysis also confirms that ultimately a more comprehensive collection of conditioning variables is needed in order to capture the whole range of transition probabilities across countries because the negative impact of less well designed institutions can be mitigated by other measures.

One important limitation of this paper is the assumption of time-invariant transition probabilities. Obviously, if developing countries implement reforms such as those advised above, their goal is to move to a more favorable transition matrix that allows them to catch-up with leading economies. Hence, an extension of the regime-switching growth framework into this direction should be tried. For the time being, the availability of data and computational issues limit the possibilities of extension, so that the knowledge about determinants of the transition probabilities in the past seems to be a valuable, albeit first step.



## Appendix A

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

$$f(y_t|s_t = j, x_t, \theta) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left(-\frac{(y_t - x_t'\beta_j)^2}{2\sigma_j^2}\right). \quad (20)$$

Define  $\mathcal{Y}_t = (y_1, y_2, \dots, y_t)$ ,  $\mathcal{X}_t = (x_1, x_2, \dots, x_t)$  and  $\mathcal{S}_t = (s_1, s_2, \dots, s_t)$  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

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

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,  $\psi_{t-1}$  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:

$$\begin{aligned} \mathcal{L}_T = & \left[ \sum_{j=1}^Q I(s_1 = j) f(y_1 | s_1 = j, \psi_0, x_1, \theta^1) \rho_j \right] \\ & \times \prod_{t=2}^T \left\{ \sum_{j=1}^Q \sum_{i=1}^Q I(s_t = j | s_{t-1} = i) f(y_t | s_t = j, \psi_{t-1}, x_t, \theta^1) p_{ij} \right\} \end{aligned} \quad (21)$$

Taking logs of (21) and rearranging gives equation (7) in the main text. By separating  $\theta$  into  $\theta^1$  and  $\theta^2$ , it is immediately obvious that derivatives of the log-likelihood function with regards to the two parts of  $\theta$  are independent of each other, which facilitates the computational process considerably.<sup>41</sup>

## 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 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 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  $m = 1, \dots, M$  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

$$= \sum_{k=1}^K \sum_{m=1}^M C_m^{iter} \tilde{\ell}_{ck}(\mathcal{O}^k | \theta_m^{iter}) \quad (22)$$

$$\leq \sum_{k=1}^K \sum_{m=1}^M \max_{m=1, \dots, M} \tilde{\ell}_{ck}(\mathcal{O}^k | \theta_m^{iter}) \quad (23)$$

$$= \sum_{k=1}^K \sum_{m=1}^M C_m^{iter+1} \tilde{\ell}_{ck}(\mathcal{O}^k | \theta_m^{iter}) \quad (24)$$

$$(25)$$

$$\leq \sum_{k=1}^K \sum_{m=1}^M C_m^{iter+1} \tilde{\ell}_{ck}(\mathcal{O}^k | \theta_m^{iter+1}) \quad (26)$$

Equation (22) 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 (23)). In the classification step all countries are assigned to the cluster that maximizes  $\tilde{\ell}_{ck}$  (equation (24)). 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

---

<sup>41</sup> 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).

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 (22) - (24) continue to hold because these steps require comparisons of the different  $\hat{\ell}_{ck}$  values for one country only. The crucial question therefore becomes whether the EM algorithm maintains its monotonicity when linear parameter restriction are involved. Kim and Taylor (1995) prove that this is the case. Hence, it follows that (26) 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 Kim and Taylor (1995).

## Appendix C

Table 12: 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
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

Country	Cluster	Crisis	Stagnation	Stable Growth	Miracle Growth	Average Growth Rate
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  $(1/T) \sum_{t=1}^T P(st = j|\psi_T)$ , 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.

## Appendix D

In this section the indicators for the quality of institutions are described in more detail.

**Government Antidiversion Index:** 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 and has been introduced by Hall and Jones (1999).

**Voice and Accountability Indicator:** 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 ranges between  $-2.5$  and  $2.5$ , whereby freedom is more restricted the lower the reported indicator. It is taken from Kaufmann et al. (2008). The earliest available indicator refers to 1996.

**Independence of Judiciary:** 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 judicial independence. La Porta et al. (2004) use constitutional data for the tenure of judges and a case law dummy referring to periods before 1973. Therefore, it is not entirely clear which time period the indicator refers to.

**Business Environmental Risk Intelligence:** 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 indicator has been introduced into the literature by Knack and Keefer (1995). The earliest available indicator refers to the year 1972. The small number of developing countries considered in this index points at a potential sample selection bias.

## Appendix E

In this section the descriptive statistics are reproduced if those countries that had already been well developed in 1960 are 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.

Table 13: Factor Accumulation - Alternative Classification

	Investment Rate 1960	Saving Rate 1965	Population Growth Rate 1960	Dependency Ratio 1960	Average Years of Schooling 1960	Primary School Com- pleted 1960
Cluster 1	17.15 <sup>4</sup> (2.40) 18	20.92 (2.78) 14	2.47 <sup>4</sup> (0.36) 18	0.80 <sup>3,4</sup> (0.04) 17	3.77 <sup>2,3,4</sup> (0.43) 18	39.36 <sup>2,3,4</sup> (4.11) 18
Cluster 2	11.97 <sup>4</sup> (1.57) 22	14.23 <sup>4</sup> (1.83) 18	2.48 <sup>4</sup> (0.15) 23	0.85 <sup>4</sup> (0.02) 23	2.50 <sup>3,4</sup> (0.41) 23	25.88 <sup>3,4</sup> (3.69) 21
Cluster 3	12.71 <sup>4</sup> (3.14) 28	14.63 <sup>4</sup> (2.47) 23	2.53 <sup>4</sup> (0.12) 28	0.89 <sup>4</sup> (0.02) 28	1.43 <sup>4</sup> (0.24) 28	13.73 <sup>4</sup> (1.68) 25
Cluster 4	24.90 (1.38) 15	26.51 (1.77) 8	1.17 (0.17) 15	0.59 (0.02) 15	7.81 (0.34) 15	75.57 (2.86) 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 Satterthwaite's degree of freedom adjustment. The investment and population growth rates are taken from the Penn World Tables 6.2 (Heston et al., 2006), the saving rates and the dependency ratios are based on the World Development Indicators 2007 (WorldBank, 2007). The average years of schooling in 1960 refer to the non-studying population aged 15-64 and are taken from Cohen and Soto (2001). The percentage of the population having completed at least primary school is taken from Barro and Lee (1993, 2001). Linear regression methods and the additional data set provided by Lutz et al. (2007) are used in order to extend the data coverage of the latter two datasets as much as possible.

Table 14: Structural Features

	Private Sector Credit to GDP 1965	Openness Dummy 1970	Years of Liberal- ized Trade 1960 – 2002	Landlocked Country	Percentage of Land Within 100km to the Coast	Percentage of Land in Tropical Climate
Cluster 1	29.33 <sup>3</sup> (5.78) 15	0.61 <sup>2,3,4</sup> (0.12) 18	28.61 <sup>2,3,4</sup> (3.40) 18	0.06 <sup>3</sup> (0.06) 18	0.59 <sup>3</sup> (0.09) 18	0.50 <sup>3,4</sup> (0.11) 18
Cluster 2	16.68 <sup>4</sup> (2.07) 20	0.13 <sup>4</sup> (0.07) 23	15.04 <sup>3,4</sup> (1.90) 23	0.13 (0.07) 23	0.51 <sup>3</sup> (0.08) 23	0.58 <sup>3,4</sup> (0.10) 23
Cluster 3	11.93 <sup>4</sup> (1.34) 23	0.04 <sup>4</sup> (0.04) 28	7.61 <sup>4</sup> (1.46) 28	0.36 (0.09) 28	0.14 <sup>4</sup> (0.03) 28	0.85 <sup>4</sup> (0.06) 28
Cluster 4	44.30 (6.54) 15	0.93 (0.07) 15	40.00 (1.73) 15	0.13 (0.09) 15	0.45 (0.09) 15	0.03 (0.03) 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 Satterthwaite's degree of freedom adjustment. Credit to GDP is taken from the World Development Indicators 2007 (WorldBank, 2007). The openness indicators are taken from Sachs and Warner (1995) and Wacziarg and Welch (2008), the geographic variables from Gallup et al. (1999).

Table 15: Quality of Institutions

	Government Antidiver- sion Index (GADP)	Voice and Account- ability (VA1996)	Independence of Judicity	BERI index 1972	BERI index 1995
Cluster 1	0.69 <sup>2,3,4</sup> (0.04) 17	0.28 <sup>3,4</sup> (0.17) 18	0.69 (0.08) 13	10.51 <sup>2,3,4</sup> (0.54) 12	10.07 <sup>2,3</sup> (0.60) 13
Cluster 2	0.53 <sup>4</sup> (0.02) 22	-0.09 <sup>3,4</sup> (0.14) 23	0.79 (0.08) 13	6.74 <sup>4</sup> (0.74) 11	7.71 <sup>4</sup> (0.43) 8
Cluster 3	0.49 <sup>4</sup> (0.01) 28	-0.54 <sup>4</sup> (0.14) 28	0.73 (0.09) 15	7.22 <sup>4</sup> (0.55) 5	6.84 <sup>4</sup> (0.58) 5
Cluster 4	0.96 (0.01) 15	1.32 (0.05) 15	0.84 (0.06) 15	12.56 (0.37) 10	11.13 (0.29) 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 Sattersthaite's degree of freedom adjustment. The data is taken from Hall and Jones (1999), Kaufmann et al. (2008), La Porta et al. (2004) and Knack and Keefer (1995).

## Appendix F

A multinomial logit model is estimated.<sup>42</sup> A major problem of the estimation the small number of observations overall and in each cluster, which precludes the use of the general-to-specific approach of model specification to arrive at a good estimation model. Rather, a sequential approach similar to Berg et al. (2008) is followed, i. e. a parsimonious model including the most important control variables is estimated and sequentially extended. This procedure increases the danger of omitting relevant variables, which in non-linear regression models lead to biased and inconsistent regressors. Fortunately, some recent contributions on binary logit models show that in practise omitted variables in logit models are not a big problem. While the coefficient estimates are indeed subject to an attenuation bias, prediction remains accurate and Wald tests of significance remain valid. The estimated parameters can be interpreted by calculating the average marginal effects over the whole sample because then the attenuation bias inherent in the parameter estimate is mitigated by an inverse change of the probability density function, rendering the average marginal effect unbiased (Cramer, 2007; Ramalho and Ramalho, 2007). 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. Therefore, Table 11 in the main text not only includes the coefficient estimates and the corresponding standard errors, but also the average marginal effects. Since the multinomial logit model is estimated via maximum likelihood, the calculated standard errors are valid only asymptotically. Given the small number of observations in the actual estimation, all results should only be interpreted as suggestive.

In any growth regression, causality is an important problem. Unfortunately, due to the small number of observations, it is not possible to use control function strategies in order to test the significance of potentially endogenous variables such as the number of years a country has been liberalized to trade (Wooldridge, 2002, chap.15.7.2, and Imbens and Wooldridge, 2007). Therefore,

<sup>42</sup> At first sight, it seems tempting to fit an ordered logit model because there seems to be a natural order in the clusters: cluster one provides better growth outcomes than cluster two and cluster two is preferred to cluster three. However, in order to apply the ordered logit model a *monotonic* relationship between an underlying index and the observed outcomes has to exist (Verbeek, 2004, chap. 7.2.1). Since there is no utility theory underlying the model, it is uncertain whether this assumption is warranted and hence the multinomial logit model is the safer choice.

the analysis is restricted to the influence of initial conditions. Since the quality of institutions, too, is used in the analysis, this effectively amounts to assuming that it has not changed during the sample period or at least that the relative quality of institutions has remained unchanged between the countries. Note that compared to Jerzmanowski (2006) the potential endogeneity problems are delayed to a stage where they do no impact on the results of the growth model itself.

## Appendix G

Table 16: Correlation Matrices

	Years of School- ing 1960	Openness 1970	GADP	Coastal Land	VA 1996
Cluster 1					
Years of Schooling 1960	1.000				
Openness 1970	0.481	1.000			
GADP	0.839	0.677	1.000		
Coastal Land	-0.089	0.236	0.014	1.000	
VA 1996	0.783	0.535	0.756	0.022	1.000
Cluster 2					
Years of Schooling 1960	1.000				
Openness 1970	0.363	1.000			
GADP	0.615	0.350	1.000		
Coastal Land	0.566	0.091	0.109	1.000	
VA 1996	0.825	0.331	0.597	0.423	1.000
Cluster 3					
Years of Schooling 1960	1.000				
Openness 1970	0.363	1.000			
GADP	0.330	0.201	1.000		
Coastal Land	0.230	0.273	0.277	1.000	
VA 1996	0.332	0.186	-0.062	0.439	1.000

Years of Schooling 1960 denotes the average years of schooling in the population in 1960, openness 1970 is a dummy variable that take on the value 1 if a country has been open in 1970, GADP is the government antidiversion index, coastal land stands for the percentage of land within 100km to the coast and VA 1996 is the voice and accountability indicator.

## Appendix H

In this section the multinomial logit results for the alternative country classification (separate cluster for the industrialized countries of 1960) are presented. The results support the conclusion drawn in the main text. The quality of institutions continues to be a highly significant predictor for both clusters one and four. If the GADP indicator is replaced by the VA indicator, institutions are still measurably better in cluster four, but no longer in cluster one compared to the reference group cluster three. As in the main text, now education turns significant in both clusters. The main difference between moderately successful countries in cluster two and the reference groups continues to be education and geography. Unlike in the main text, openness in 1970 is never significant. However, the percentages correctly predicted improve markedly from column five to column six in cluster one, so that openness appears to be importance nevertheless.



Table 17: Multinomial Logit Specifications

	(1)	(2)	(3)	(4)	(5)	(6)
Cluster 1						
Years of schooling 1960	0.389 (0.291)	0.449 (0.284)	0.232 (0.308)	0.704** (0.284)	0.535* (0.305)	0.514 (0.329)
Openness Dummy 1970	0.007 (1.311)	0.012	0.011	0.042	0.031	0.020 (1.282)
GADP	0.197					0.309
	10.927** (4.715)	13.097** (4.469)	13.644** (4.994)			
Coastal Land	0.217	0.507	0.530 3.300* (1.702)		3.657** (1.610)	2.802 (1.712)
VA 1996			0.148	0.610 (0.633)	0.246 (0.680)	0.246 (0.680)
				-0.078	-0.066	-0.088
Cluster 2						
Years of schooling 1960	0.480** (0.227)	0.520** (0.226)	0.199 (0.267)	0.433* (0.251)	0.212 (0.286)	0.213 (0.286)
Openness Dummy 1970	0.057 (1.257)	0.060	0.015	0.018	-0.015	-0.003 (1.398)
GADP	-0.019					-0.225
	2.186 (3.509)	2.188 (3.370)	3.109 (4.103)			
Coastal Land	-0.265	-0.419	-0.413 4.160** (1.426)		4.448** (1.543)	4.542** (1.614)
VA 1996			0.401	0.394 (0.498)	-0.071 (0.584)	-0.071 (0.584)
				-0.031	-0.065	-0.029
Cluster 4						
Years of schooling 1960	0.498 (0.622)	0.552 (0.606)	0.209 (0.778)	1.377** (0.480)	1.478 (0.688)	1.457** (0.677)
Openness Dummy 1970	0.002 (4.397)	0.002	-0.000	0.022	0.027	0.024 (2.517)
GADP	-0.001					0.053
	40.150** (17.344)	42.172** (16.796)	46.460** (21.542)			
Coastal Land	0.680	-0.683	-0.657 -0.964 (3.969)		0.367 (3.395)	-0.535 (3.597)
VA 1996			-0.083	7.322** (3.278)	5.890 (3.300)	4.926 (3.180)
				0.192	0.148	0.116
Pseudo $R^2$	0.454	0.440	0.506	0.373	0.435	0.477
$\chi^2$	100.990	97.870	112.310	83.050	96.700	106.220
p-value	0.000	0.000	0.000	0.000	0.000	0.000

	(1)	(2)	(3)	(4)	(5)	(6)
Proportion correctly predicted						
Cluster 1	0.588	0.647	0.647	0.176	0.235	0.411
Cluster 2	0.455	0.318	0.545	0.227	0.455	0.545
Cluster 3	0.821	0.821	0.857	0.786	0.857	0.857
Cluster 4	0.933	0.933	0.933	0.867	0.933	0.933

Standard errors are reported in parenthesis. \* denotes significance at the 10%-level, \*\* significance at the 5%-level. The third line for each variable reports the average partial effect. The number of observations in total is 82 (81 in (3) due to Tunisia). Cluster one contains 17, cluster two 22, cluster three 28 , cluster four 15 observations.

## 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.
- Aghion, P. and S. Durlauf (2005). *Handbook of Economic Growth*. Elsevier.
- Aizenman, J. and M. Spiegel (2007). Takeoffs. *NBER Working Paper Series 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. J. and J.-w. Lee (1993). International Comparisons of Educational Attainment. *Journal of Monetary Economics* 32, 363–394.
- 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. *Tinbergen Institute Discussion Paper 2008-085/4*.
- Becker, T. and P. Mauro (2006). Output Drops, and the Shocks that Matter. *IMF unofficial working paper*.
- Ben-David, D. (1998). Convergence Clubs and Subsistence Economies. *Journal of Development Economics* 55, 155–171.
- Berg, A., C. Leite, and J. D. a. Ostry (2008). What Makes Growth Sustained? *International Monetary Fund Working Paper*.
- Berthelemy, J.-C. (2006). Convergence and Development Traps: How Did Emerging Economies Escape the Underdevelopment Trap? *Annual World Bank Conference on Development Economics 2006*.
- 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.
- Bos, J. W. B., C. Economidou, and M. Koetter (2010). Do All Countries Grow Alike? *Journal of Development Economics* 91, 113–127.
- 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. *Brown University Working Paper 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). Identifying US Turning Points Revisited: The Panel Model with the Regime Switching Approach. *Applied Economics Letters*, 1–5.
- Chen, S.-W. (2007b). Using Regional Cycles to Measure National Business Cycles in the U.S. with the Markov Switching Panel Model. *Economic Bulletin* 3, 1–12.
- Cohen, D. and M. Soto (2001). Growth and Human Capital: Good Data, Good Results. *CERP Discussion Paper 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.
- 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. Titterton (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. 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. *KSG Working Paper RWP06-046*.
- Heston, A., R. Summers, and B. Aten (2006). Penn World Table Version 6.2. *Center for International Comparisons of Production*.
- Imbens, G. W. and J. M. Wooldridge (2007). Control Function and Related Methods. *What's New in Econometrics? NBER Summer Institute 2007*.
- 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. *CESifo Working Paper 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. *World Bank Policy Research Working Paper 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.
- Kim, D. K. and J. M. G. Taylor (1995). The Restricted EM Algorithm for Maximum Likelihood Estimation Under Linear Restrictions on the Parameters. *Journal of the American Statistical Association* 90, 708–716.
- 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*. Dissertation Mathematisch-Naturwissenschaftliche Fakultät Köln, <http://deposit.ddb.de/cgi-bin/dokserv?idn=96144195X>.
- 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 E. S. Prasad (2008). Global Business Cycles: Convergence or Decoupling? *IZA Working Paper 3442*.
- 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 Approach to Cross Country Growth Data. *Department of Economics, University of Cyprus Discussion Paper 2002-13*.
- Krolzig, H.-M. (1997). *Markov Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*. Springer.
- La Porta, R., F. Lopez-de Silanes, and C. Pop-Eleches (2004). Judicial Checks and Balances. *Journal of Political Economy* 112, 445–470.
- Li, C. and G. Biswas (2000). A Bayesian Approach to Temporal Data Clustering Using Hidden Markov Models. *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 543-550.
- 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. *Vienne Yearbook of Population Reserach, Vol. 2007*, pp. 193-235.
- 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 Non-linear Models? A Simulation Exercise for Binary and Fractional Data. *CEFAGE-UE Working Paper 2007/08*.

- Rodrik, D. (2003). *In Search of Prosperity: Analytic Narratives on Economic Growth*. Princeton University Press.
- Rodrik, D. (2005). Growth Strategies. See Aghion and Durlauf (2005), pp. 967–1014.
- Sachs, J. D. and A. Warner (1995). Economic Reform and the Process of Global Integration. *Brookings Papers on Economic Activity 1995*, 1–118.
- 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.
- 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.