The Gap Between the Rich and the Poor: Patterns of Heterogeneity in the Cross-Country Data

Oya Pınar Ardıç*†

April 1, 2005

Abstract

Persistent income inequalities across nations have led to the emergence of a voluminous literature, focusing on cross-country growth regressions. Using the same regression model for all countries in the sample, the majority of these studies ignore the inherent heterogeneity that can actually lead to different regression models for different countries. This paper explores whether this assessment is valid, and in doing so, provides a way to overcome it. Bayesian classification analysis is used to reveal patterns of heterogeneity and to identify groups of countries with similar growth processes. Standard growth regressions can then justifiably be performed on each subsample. The method is illustrated using a cross-country data set that includes the Solow growth model variables.

Keywords: Cross-country growth; Heterogeneity; Bayesian classification
JEL Classification: O47; C21; C11

*Bogazici University, Department of Economics, Bebek 34342, Istanbul, Turkey. Tel: +90(212)359-7650 Fax: +90(212)287-2453 e-mail: pinar.ardic@boun.edu.tr URL: http://www.econ.boun.edu.tr/ardic

†I thank Steven Durlauf, William Brock, and Louise Keely for support, advice and encouragement, Faruk Selçuk for discussions, and seminar participants at the University of Wisconsin - Madison and an anonymous referee for useful comments. All errors are mine.

1 Introduction

The real GDP per capita, the widely used measure of the standard of living, in the Democratic Republic of Congo was $216 whereas it was $19,474 in the United States in 1996. This striking fact exemplifies the gap between the rich and the poor, and suggests that the richest are about 90 times richer than the poorest. Furthermore, there is evidence that this gap is widening: in 1977, the real GDP per capita in the Democratic Republic of Congo was $557 while it was $14,832 in the US, which indicates a 27 times difference. Why is there a huge gap between the rich and the poor? How can the poor catch up with the rich?

The gap between the rich and the poor nations has been a perennial concern of economists. In the face of persisting income inequalities across nations, a voluminous literature has emerged, mainly focusing on cross-country growth regressions to investigate the relationship between the per capita income growth and a set of other indicators of country characteristics. Though these empirical models have improved our understanding of the mechanics of economic growth significantly, they have not been free from serious criticisms. An important criticism is that the majority of these studies treat countries which have intrinsic differences as homogeneous units, using the same regression model for all countries in the sample, thereby ignoring the high degree of heterogeneity in the cross-country growth data. Therefore, empirical methods that allow for heterogeneity might yield substantially different findings regarding the important determinants of growth.[1]

The purpose of this study is to reconsider the empirical growth models by exploring the uncertainty for heterogeneity in the cross-country growth data. Although some studies have attempted to resolve this problem in a systematic way, the general practice is to ignore it. Using fixed effects in panel data regressions, or including dummy variables based on geographic locations or informal groupings of the data, will neither reveal the underlying patterns of heterogeneity nor provide an adequate solution to the problem, since such ad hoc adjustments assume a particular pattern of heterogeneity.

[1] See Durlauf (2000) and Brock and Durlauf (2001) for details.

The paper proposes the use of Bayesian classification analysis (or cluster analysis) in order to first systematically reveal the patterns of heterogeneity in the data set. The objective of classification analysis is to partition the data into subsamples that display systematic differences, and the Bayesian method aims to find the classification that fits the data with the highest probability. The second step, then, is to use this information on heterogeneity in the cross-country growth regressions and in the tests of (cross-country) convergence hypothesis by performing a separate regression analysis for each subsample.

To illustrate, an exemplifying cross-country data sets are used, which includes the standard Solow growth model variables (Solow, 1956). This data set have also been used by Mankiw *et al.* (1992) (MRW hereafter), which is an essential study in the empirical growth literature. The findings of the analysis on this data set indicate that once the heterogeneity in the sample is accounted for according to the underlying statistical distributions, the regression outcomes differ from what the Solow model predicts. Thus, this two-step method helps us better understand the properties of the sample, i.e. the patterns of heterogeneity, and of the cross-country growth process, i.e. the possibility of multiple regimes.

The rest of the paper is organized as follows. The next section briefly notes the existing empirical studies. The third section describes the Bayesian classification method. The results of MRW model are analyzed and reported in the fourth section. The fifth section concludes.

2 Existing Empirical Studies

In the 1990s, there have been various studies on the relationship between the growth rate of income per capita and different measures of standard of living in a cross-country setting to investigate the growth process. These studies focus on a model of the form:

$$g_i = \alpha X_i + \beta y_{i0} + \epsilon_i \tag{1}$$

where g_i is the growth rate of country level variable, y_{i0} is the value of the country level variable at the beginning of the period of analysis, X_i includes country-specific variables that are controlled for, and ϵ_i is the disturbance term. The initial value of the variable, y_{i0}, is included for the purpose of testing for the convergence hypothesis (Durlauf, 2000).[2]

The convergence hypothesis states that the poor countries tend to grow faster than the rich due to diminishing marginal returns, since the returns to capital would be higher in those countries with lower initial conditions. One of the convergence concepts commonly used in the literature is β-convergence. There exists a β-convergence across countries if there is a negative relationship between the per capita income growth rate and the initial value of per capita income. That implies poor countries grow faster than rich ones. In terms of equation (1), β-convergence means a negative β when g_i is the growth rate of per capita income and y_{i0} is the initial value of per capita income in country i. If the country-specific controls, X_i, are not used in the analysis, a negative relationship between the growth rate and the initial value implies unconditional (or absolute) convergence, whereas it indicates conditional convergence when controls are included (Barro and Sala-i Martin, 1995). Therefore, equation (1) facilitates the tests of convergence hypothesis.

Note however that there are also studies that criticize this approach of testing for convergence. For example, Bernard and Durlauf (1996) state that once this analysis is applied to a data set of countries that can be correctly specified with a model with multiple steady states, an estimated negative β coefficient implying convergence for the whole sample can actually arise from within-subsample convergence to group-specific steady states. In addition, Quah (1993, 1996b) suggests the tests of convergence hypothesis suffer from Galton's fallacy, i.e. once the average growth rates are regressed on initial levels, a negative β coefficient is estimated due to regression toward the mean, which does not necessarily imply convergence.

[2]Generally the growth rate of income per capita is used, however it is possible to use the growth rate of any standard of living indicator.

The theory of growth is not clear on the true set of explanatory variables to be included in the growth regression, leaving the question of which variables can explain the growth process unanswered. Various measures including investment rate, education, policy indicators among many others have been found to explain the growth rate of different indicators by researchers.[3] For example, Barro (1991) reports the empirical regularities about growth, education, fertility and investment in cross-country data, and finds evidence supportive of convergence across countries. Mankiw *et al.* (1992) provide an empirical analysis of the Solow model with a production function with human capital, physical capital and labor as factors, and conclude that there exists conditional convergence among the countries in their data set. Further examples include Barro and Lee (1993) who consider the relationship between income growth and education, Mauro (1995) who focuses on corruption and growth, and Barro (1996) who investigates the relationship between democracy and income growth among numerous other studies.[4] Overall, around 90 different standard of living indicators have been used in this literature.

Levine and Renelt (1992) propose the use of extreme bounds analysis to find the robust explanatory variables, and conclude that there are few. Sala-i Martin (1997) computes the distribution of the coefficient estimates in equation (1) and uses confidence levels based on these distributions to find variables that are correlated with growth rather than labeling variables as robust or non-robust. Easterly (1999) provides an extensive investigation of the relationship between the quality of life and income per capita, concentrating on a variety of indicators of quality of life to observe the ones that are related to growth of income per capita. Brock and Durlauf (2001) allow for uncertainty in model specification, and use Bayesian techniques to determine the explanatory variables.

All of the studies mentioned above, with the exception of Brock and Durlauf (2001), essentially ignore the underlying patterns of heterogeneity in the data, by imposing an identical regression model for all countries in the sample. Some of them use dummy

[3]See, for example, Sala-i Martin (1997).

[4]See Durlauf and Quah (1999) for an extensive review of the literature.

variables for Latin America and/or Sub-Saharan Africa to account for the differences in growth processes for these groups of countries, however this does not capture the statistical groupings in the data set.[5] As a result, these studies estimate identical parameters for each country. To put this into the perspective of the theory that underlies these empirical studies, these estimated identical parameters imply that the production function for each country in the data set is identical, and the growth processes of each are modeled the same way.[6] On the other hand, it is natural to expect that the countries at different levels of development have production functions with different parameters, and exhibit different growth processes.

Studies that incorporate a systematic form of heterogeneity include Canova (1999) who proposes the use of a predictive density approach to jointly test for the groupings of unknown size and estimate the parameters for each group in identifying convergence clubs and applies this to European and OECD data, and various studies by Quah (1996a, 1997) who adopts the distribution dynamics approach and concludes that the cross-country data supports the twin-peaks hypothesis. In addition, Durlauf and Johnson (1995) use regression tree analysis and they find evidence supporting the presence of multiple regimes in cross-country growth data in which each group of countries follows different linear models, namely growth models that produce multiple steady states in per capita output. Kourtellos (2001) uses projection pursuit regression and finds cross-country evidence supporting two equilibria with different convergence parameters. Brock and Durlauf (2001) propose modeling heterogeneity as a form of model uncertainty using Bayesian techniques. Durlauf *et al.* (2001) use the Solow growth model, and allow the parameters to differ across countries according to initial income.

Hobijn and Franses (2001) study the cross-country convergence problem using three different techniques: regression analysis, distribution dynamics, and (classical) classifi-

[5]For example, Barro (1991) uses dummy variables for Sub-Saharan Africa and Latin America.

[6]The baseline regressions of the form (1) can be derived from the Solow model, and the coefficient estimates in these regressions are the parameter estimates of the production function. See Mankiw *et al.* (1992).

cation analysis.[7] The major question the paper proposes is whether there exists similar convergence patterns for standard of living indicators other than income per capita to those observed for income per capita. Hobijn and Franses (2001) use equation (1) and subsamples from their data set based on the World Bank country classification to test for β-convergence in each subsample, and find evidence indicating a possible convergence in the tails of the distribution in terms of income per capita. They utilize kernel density estimation for the distribution dynamics approach and again find a similar result. These results for real GDP per capita also hold for the other indicators in their data set. Finally, they apply (classical) classification analysis to each variable in the separately. Their main conclusion is that there is not much evidence of convergence in any of the indicators in their data set, and that convergence in one indicator does not imply convergence in another.

The analysis of this paper differs from Hobijn and Franses (2001) in crucial respects. First note that, although being helpful in identifying the presence of multi-modality in the data set, kernel density estimation cannot provide enough information about the specific groupings of countries in the data set.[8] Second, applying classification analysis on the whole set of standard of living indicators rather than examining each variable separately yields insight as to how countries are clustered in terms of the level of development measured by a variety of indicators. Third, once the countries are grouped into clusters that display systematic differences, it is possible to discuss the existence of within-cluster convergence. Fourth, Bayesian classification analysis has advantages over the classical classification analysis as mentioned in Section 3 below.

[7] See Hobijn and Franses (2000) for the method of classification used in Hobijn and Franses (2001). where the authors develop a cluster algorithm and apply it on the cross-country per capita productivity levels.

[8] In addition, Bayesian classification analysis also is useful in studying the dynamics of the distribution of standard of living across countries when it is applied to a cross section of countries at different periods of time. See, for example, Ardıç (2004).

3 Bayesian Classification Based on Finite Mixture Models

Classifying a set of data, i.e. arranging data into groups of similar nature, can be supervised or unsupervised. Supervised classification refers to grouping objects into given labeled clusters. The predefined classes are differentiated by criteria that maximizes in-class similarity and out-class dissimilarity. In unsupervised classification, there are no preexisting clusters, and all features of new observations are predicted. The goal is to discover natural classes that arise from the underlying mechanisms, to divide the data into groups that display systematic differences. In the current context, the aim is to classify the countries, based on their standard of living indicators, into groups to identify countries with similar growth processes.

The classical approach to classification analysis aims to maximize between-cluster variation relative to within-cluster variation. The clustering procedure starts with K measures of I objects, and thus, an $I \times K$ data matrix. This matrix is then transformed to an $I \times I$ matrix of pairwise similarities or dissimilarities. Finally, an algorithm that defines the rules of classifying the objects into subgroups is selected (Dillon and Goldstein, 1984).

The problems involved with the classical approach include sensitivity to the variables (or to the features of data) used in the analysis, the definition and measurement of similarity (and dissimilarity), and deciding on the procedure and the number of clusters (Dillon and Goldstein, 1984). In most cases, the researcher needs to use some ad hoc criteria to decide on the number of clusters. The classical approach lacks a widely accepted measure of success, and the method favors singleton clusters.[9] In addition, small changes in the decision criteria might alter all the results by changing the clusters that the boundary cases belong to (Stutz and Cheeseman, 1996).

The Bayesian approach to unsupervised classification aims to find the classification that fits the data with the highest probability. This procedure looks for the natural classes in the data. The outcome is the probabilities of having certain numbers of classes in the data set. Furthermore, instead of assigning each case to a class, the Bayesian

[9]For example, Hobijn and Franses (2001) find a large number of clusters.

approach yields the probabilities of each case being members of different classes.[10] This helps overcome the problems related to the decision criteria; the boundary cases are no longer a problem. The number of classes are determined according to the probability assignments, i.e. the classification with the highest probability is chosen. Note that it is also possible to rank the alternative classifications with this approach (Stutz and Cheeseman, 1996). Furthermore, the Bayesian approach trades off complexity for goodness of fit. However, it is important to note that Bayesian classification is also sensitive to the variables included in the analysis.

Bayesian approach to statistics in general enables us to express all forms of uncertainty in terms of probability. Bayesian theory explains how beliefs should be formulated in a consistent way and how they should change with new evidence. Let Y denote evidence (or data) and θ denote the parameters of the model that we are interested in. Also let $P(\theta)$ denote the prior belief in θ before the data Y is observed, $P(Y \mid \theta)$ be the likelihood of the data for each possible θ, and $P(Y, \theta)$ be the joint distribution of Y and θ. Then, this joint distribution can be obtained from the observable likelihood, $P(Y \mid \theta)$, and the assumed prior, $P(\theta)$. That is, it is possible to think of the joint probability distribution of the data, Y, and the parameters, θ, as the multiplication of the likelihood of the data and the probability distribution summarizing the prior belief over the different possible values that the parameters, θ, can take.

The joint distribution, $P(Y, \theta)$, should remain the same once the parameters, θ, are inferred from the data. Therefore, it is possible to consider the joint distribution as a multiplication of the probability distribution of the parameters given the data set, $P(\theta \mid Y)$, which is called the posterior distribution of the parameters, and the prior predictive distribution, $P(Y)$. $P(Y)$ is called the prior predictive distribution because it is the distribution of an observable quantity that does not depend on any previously observed values, that is, it is not derived. Given these, the posterior belief in θ, $P(\theta \mid Y)$, becomes a function of the likelihood, the prior, and the prior predictive distribution. In

[10] This is also termed as "fuzzy" classification.

other words, the probability distribution of the parameters, θ, given the evidence, Y, is found by the likelihood times the prior belief, normalized by the prior predictive distribution, the distribution of the evidence that does not depend on the parameters.

The advantages of using Bayesian analysis include a good theoretical basis, the possibility of using background knowledge as an input, and getting output in terms of probabilities rather than a definite answer. Most of the disadvantages put forth are in terms of the ambiguities involved in choosing a prior, which are not important in practice since a broad range of priors are found to perform well under most situations (Hanson *et al.*, 1991).

A Bayesian approach based on finite mixture distributions is the recent focus of a study of the Bayes group at the Ames Research Center.[11] The group has been working on a software called *AutoClass* that utilizes the Bayesian approach to unsupervised classification. Their method is applied in this paper, and the remainder of this section closely follows their treatment.

Mixture distributions arise when one samples from a heterogeneous population. If the number of subcomponents of the population is finite, then it is a finite mixture distribution. Mixture modeling provides a natural way to handle the unobserved heterogeneity in the data set since the aim is to model the distribution of the whole sample as a mixture, or weighted sum, of the distribution of subsamples.

In terms of exploring the heterogeneity in the cross-country data set, the mixture model can be motivated as follows. The data set is a mixture distribution with J components, that is, the sample has J subgroups, J is unknown. Within each subgroup, the countries have the same growth process, i.e. the same regression model can be applied to subsamples as the data for the countries in a subsample have the same statistical distribution. Thus, Bayesian classification in this setting would allow us to identify J, the number of subsamples, and the countries belonging to each subsample. Below is a formal description of Bayesian classification based on finite mixture distributions.

[11] See http://ic.arc.nasa.gov/ic/projects/bayes-group/autoclass/, and Stutz and Cheeseman (1996) and Hanson *et al.* (1991).

Let $y = \{y_1, ..., y_I\}$ be the data set where $i = 1, ..., I$ shows the observations. y is sampled from a heterogeneous population, the components of which are indexed by $j = 1, ..., J$. Let $k = 1, ..., K$ be the attributes. Thus, each y_i is a $(1 \times K)$ vector, and y is an $(I \times K)$ matrix of data sampled from a population of J components.

Further, let $F = F_1, ..., F_J$ denote the mathematical form of the probability distribution function associated with J components, and $\theta = \theta_1, ..., \theta_J$ be the parameter set for each of the J distributions. That is, for each class that is to be identified, there is a distribution function for the attributes, F_j, with parameters θ_j.

Let T denote the inter-class mixture model. F_j is weighted by a mixture model T, i.e. the probability distribution that any y_i is a member of class j, C_j, regardless of its attribute values. The parameters of T are π_j which are defined as follows. The proportion of the population that is from component j is given by π_j, $\sum_{j=1}^{J} \pi_j = 1$, and $\pi = (\pi_1, ..., \pi_J)$. Then it is possible to write the likelihood of the observation y_i as:

$$\begin{aligned} P(y_i|\theta, \pi) &= \pi_1 P(y_i|\theta_1, F_1) + ... + \pi_J P(y_i|\theta_J, F_J) \\ &= \sum_{j=1}^{J} \pi_j P(y_i|y_i \in C_j, \theta_j, F_j) \end{aligned} \tag{2}$$

where $\pi_j = P(y_i \in C_j \mid \pi, T)$ as described above. $\pi = \{\pi_1, ..., \pi_J\}$ can be taken as a mixing distribution that describes the variation of θ_j across the population. Thus, $\nu = (\theta, \pi)$ is the set of parameters of the model. Let $\mathcal{M} = (F, T)$, and $\mathcal{M} \in S$ where S is the space of possible mixture models. Then the likelihood of the whole sample is given by:

$$P(y|\nu, \mathcal{M}) = \prod_i \sum_j \pi_j P(y_i|y_i \in C_j, \theta_j, F_j) \tag{3}$$

The joint distribution of the data and the parameters can be written as prior times the likelihood:

$$\begin{aligned} P(y, \nu|\mathcal{M}) &= P(\nu|\mathcal{M}) P(y|\nu, \mathcal{M}) \\ &= P(\nu|\mathcal{M}) \prod_i \sum_j \pi_j P(y_i|y_i \in C_j, \theta_j, F_j) \end{aligned} \tag{4}$$

where the prior can be expressed as:

$$P(\nu|\mathcal{M}) = P(\pi|T)P(\theta|F) \tag{5}$$

since π and θ are independent.

The objective is to find the posterior distribution of the parameters, and the MAP (maximum a posteriori) values of the parameters. The posterior distribution is:

$$P(\nu|y,\mathcal{M}) = \frac{P(y,\nu|\mathcal{M})}{P(y|\mathcal{M})} = \frac{P(y,\nu|\mathcal{M})}{\int P(y,\nu|\mathcal{M})\mathrm{d}\nu} \tag{6}$$

In addition, the posterior probability of the model given the data is also calculated, in order to enable comparisons of alternative classifications:

$$\begin{aligned} P(\mathcal{M}|y) = \quad & P(\mathcal{M},y)/P(y) \\ & = \left[\int P(y,\nu|\mathcal{M})P(\mathcal{M})\mathrm{d}\nu\right]/P(y) \\ & \propto \int P(y,\nu|\mathcal{M})\mathrm{d}\nu = P(y|\mathcal{M}) \end{aligned} \tag{7}$$

where the proportionality in equation (7) holds if $P(\mathcal{M})$ is assumed to be uniform. This is sensible since there is no reason to favor one model over another. $P(\mathcal{M}|y)$ provides a measure of how well the possible classification models can describe the data.

To find the MAP parameter values, direct optimization is not useful. Recall the assumption underlying the mixture models that each observation is the member of only one class. Thus, $P(y_i|y_i \in C_j, \theta_j, F_j) = 0$ whenever $y_i \notin C_j$. This enables us to eliminate the summation and rewrite equation (4) as:

$$P(y,\nu|\mathcal{M}) = P(\nu|\mathcal{M}) \prod_j \prod_{y_i \in C_j} \pi_j P(y_i|\theta_j, F_j) \tag{8}$$

Note that for the case of supervised classification where J is known, it is straightforward to maximize equation (8). However, in unsupervised classification, the number of classes is unknown, and searching for every single partitioning of the data and maximizing does not seem plausible with large data sets. In this case, the EM algorithm (Dempster *et al.*, 1977) can be used given the set of F_j and the current MAP estimates

of ν, the expectation step of the algorithm yields class assignments, ω_{ij}, in the following form:

$$\omega_{ij} = P(y_i \in C_j | \nu, \mathcal{M}) \propto \pi_j P(y_i | y_i \in C_j, \theta_j, F_j) \tag{9}$$

These weights allow the construction of statistics that can be used in the maximization step, i.e. in finding the MAP parameter values in the equation (8). Successive implementation of these two steps lead to MAP parameter values converging to a stable local maximal point. Note, however, that there is more than one such point. Thus, the *Autoclass* software searches and collects a set of such local maxima. Next, $P(y|\mathcal{M})$ is computed for each, which is used to approximate $P(\mathcal{M}|y)$ (see equation (7)), and the models are ranked according to their largest $P(y|\mathcal{M})$.

Autoclass uses Bernoulli distributions with Dirichlet prior for discrete variables. For continuous variables, Gaussian model is used with Gaussian prior for the mean and inverse-Wishart distribution for the variance. In both cases, it is possible to model the variables as independent or covariant. When a continuous variable is bounded below, i.e. for example it cannot take negative values, the log transform is taken first, and then the Gaussian model is applied. The proportions π_j have a multinomial distribution with a Dirichlet prior. Note that conjugate priors are used so that the posterior has the same form as the prior which enables its use as prior in the subsequent steps. In addition, the prior on the number of classes and the class distributions, $P(\mathcal{M})$, is taken to be uniform.

This paper proposes the use of Bayesian classification method outlined above as the first step in empirical analysis of cross-country growth. This enables to group the data into classes such that the statistical distribution of the data in each class is different. The second step is to estimate cross-country growth regressions for each class. An alternative approach is regression tree analysis used by Durlauf and Johnson (1995). Regression tree groups countries with similar linear regression models, and diminishes country-specific heterogeneity as it accounts for the possibility of multiple steady states. As Bayesian classification groups countries with similar statistical distributions, it eliminates uncertainty for heterogeneity.

4 MRW Model

This section reports the results of the analysis of the MRW Model using the proposed two-step method for two different empirical cross-country growth models. The same set of explanatory variables, which is derived from the Solow model, as in MRW is used. This section presents the Solow model briefly, and then summarizes the results of the two step analysis.

The Solow model implies that the log of income per capita, Y/L can be expressed in terms of the log of saving rate, s, and log of population growth rate, n, plus the exogenous rates of technical change, g, and depreciation, δ:

$$\ln\frac{Y}{L} = a + \frac{\alpha}{1-\alpha}\ln(s) - \frac{\alpha}{1-\alpha}\ln(n+g+\delta) + \epsilon \tag{10}$$

where ϵ is the disturbance term. When the Solow model is augmented with human capital, an equation for log income per capita similar to (10) can be derived:

$$\begin{aligned}\ln\tfrac{Y}{L} = \quad & \ln A_0 + gt - \tfrac{\alpha+\beta}{1-\alpha-\beta}\ln(n+g+\delta) \\ & + \tfrac{\alpha}{1-\alpha-\beta}\ln(s_k) + \tfrac{\beta}{1-\alpha-\beta}\ln(s_h)\end{aligned} \tag{11}$$

where s_k and s_h are the saving rates in terms of physical and human capital respectively. In addition, to test for unconditional convergence the following equation is used:

$$\ln y_t - \ln y_0 = (1-e^{-\lambda t})\ln y^* - (1-e^{-\lambda t})\ln y_0 \tag{12}$$

where y is the per capita income, and λ is the convergence rate. Note that once the determinants of the steady state are substituted in equation (12) the outcome is similar to equation (1). In the context of the Solow model and its augmented version outlined above, this substitution implies the following equation:

$$\begin{aligned}\ln y_t - \ln y_0 = \quad & (1-e^{\lambda t})\tfrac{\alpha}{1-\alpha-\beta}\ln(s_k) + (1-e^{\lambda t})\tfrac{\beta}{1-\alpha-\beta}\ln(s_h) \\ & -(1-e^{\lambda t})\tfrac{\alpha+\beta}{1-\alpha-\beta}\ln(n+g+\delta) - (1-e^{\lambda t})\ln(y_0)\end{aligned} \tag{13}$$

See Mankiw *et al.* (1992) for the derivation of these equations. Equations (10) through (13) can be estimated using cross-country data to test the validity and the predictive power

of the Solow model in a cross-country setting. In the remainder of this section, data on each subsample resulting from Bayesian classification are utilized to estimate these relationships.

As noted earlier, the same data set as in MRW is used in this section, however extending the period of analysis. The data set includes real per capita GDP, y, saving rate, s_k, population growth rate, n, and schooling, s_h for the period 1960-1995 for 105 countries.[12] Table 1 summarizes the descriptive statistics. Saving rate, schooling and population growth are in terms of the averages over the period of analysis. Schooling variable, education, is the gross secondary school enrollment ratio, i.e. it is the ratio of total enrollment in secondary school regardless of age to the population of the age group that officially corresponds to secondary school. Saving rate is defined as the ratio of investment to GDP.

Preliminary analysis of the data though the descriptive statistics implies that per capita real GDP increased in the 35-year period on average, and it is accompanied by an increase in the cross-section variation. The range of incomes also increased. There are large discrepancies between the maximum and minimum average schooling in the sample. Saving rate and population growth rate show a similar pattern, all supporting the existence of heterogeneity in the sample.

The results of the replication of the analysis of MRW with the extended data set are reported in the next subsection. Then, Bayesian classification is performed on this data set to obtain the clusters. Finally, regression analyses are performed for equations (10) through (13).

4.1 MRW Analysis - No Heterogeneity

Part of the analysis in MRW is replicated with the different data set described above. Tables 2, 3, and 4 below summarize the findings. As in MRW $g + \delta$ is taken as 0.05.

Following the analysis of MRW the sample is divided into three subgroups: non-oil

[12]Source: http://www.www.worldbank.org/research/growth/GDNdata.htm, World Bank's Global Development Network Database.

producers, intermediate, OECD countries. In oil-producing countries, a large part of the recorded GDP represents the extraction of oil and therefore leaving those countries out of the sample might yield more accurate results in terms of economic growth. The intermediate sample excludes those countries that are reported to have low-quality data, or population below 1 million, due to measurement error and the argument that the determination of income may be idiosyncratic. For OECD countries, the data is assumed to be of high-quality, and variation in omitted country-specific variables is expected to be small. The number of countries in the whole sample is 105 which includes 5 oil producers, 18 OECD countries, and 34 countries that have low-quality data or population less than 1 million. This yields 100 observations for the non-oil producers subsample, 71 observations for the intermediate subsample, and 18 observations for the OECD subsample.

The Solow growth model predicts a positive relationship between income per capita and the saving rate in terms of both human and physical capital, and a negative relationship between income per capita and $n+g+\delta$ in equations (10) and (11). The results support the findings of MRW for equation (10) as signs of the estimated coefficients are as expected, and highly significant for the non-oil and intermediate subsamples. For the OECD subsample, the estimated coefficient for the saving rate is not significant at any conventional level. These are summarized in Table 2.

The relationships implied by the Solow growth model break once the estimates of equation (11) are obtained. While the saving rate for physical capital negatively impacts income per capita for the non-oil and intermediate subsamples and statistically insignificant for the intermediate group, increases in population growth raise income per capita for these two groups. The estimated coefficients for population growth are insignificant at any conventional level for all groups.

Table 3 reports the estimation results for equation (12) and the implied rate of unconditional convergence in each subsample. The coefficient estimates are not statistically significant for the intermediate sample. The results indicate that in the non-oil and intermediate subsamples exhibit divergence (-0.4% and -1%, respectively) while the OECD

countries converge at a rate of 1.3% unconditionally. These convergence coefficients are statistically significant for the non-oil and OECD samples.

Equation (13) is estimated twice, with and without saving rate in terms of human capital. Table 4 summarizes the results, as well as the implied rates of conditional convergence. The signs of the coefficient estimates in both cases are as expected. The results for the regression without human capital show that the saving rate is statistically significant for all three subsamples, while population growth is not significant for OECD countries, and initial income is not significant for non-oil and intermediate samples. The rate of conditional convergence is 0.1%, 0.4%, and 1.1% for the non-oil, intermediate, and OECD subsamples respectively, however it is insignificant statistically for the non-oil producers sample.

The inclusion of human capital in terms of schooling increases the adjusted R^2 for all samples. Estimated coefficients for population growth are statistically insignificant for all samples although the signs are as the Solow model predicts. Schooling is insignificant for OECD countries. The estimated rates of conditional convergence are higher for this regression: 0.7%, 1%, and 1.7% respectively for the non-oil, intermediate, and OECD subsamples, and all are statistically significant.

These results imply that when the period of analysis is extended from 1960-85 to 1960-95, the sharp outcomes supportive of the Solow model in the cross-country setting deteriorate.

4.2 Classification Analysis

The results of Bayesian classification are summarized in Tables 5, 6 and 7. The classification with the highest probability groups the data into four clusters. Class 1 has 24 members, Class 2 has 22, Class 3 has 43, and Class 4 has 16 members. Table 5 lists the log posterior probabilities of a few best possible alternative classification models. These probabilities can be used to calculate relative probability of two alternative models by taking the exponential of the difference between the log posterior probabilities of the two

models, i.e. the result will yield $P(\mathcal{M}_1|y)/P(\mathcal{M}_2|y)$ which gives the number of times $\mathcal{M}_1$ is more likely than $\mathcal{M}_2$.

Table 6 lists the countries in each class. For the majority of the countries, the probability of being a member of the classes shown is larger than 0.8. The countries with membership probability less than 0.8 are as follows: Belize is in Class 3 with probability 0.79, in Class 2 with probability 0.21; Mexico is in Class 3 with probability 0.63, in Class 2 with probability 0.37; Chile is in Class 2 with probability 0.57, in Class 3 with probability 0.4; Tunisia is in Class 2 with probability 0.57, in Class 3 with probability 0.43; and Zambia is in Class 4 with probability 0.75, in Class 3 with probability 0.24. In Table 6 the countries are listed under the groups for which their membership probability is the highest.

Table 7 summarizes the descriptive statistics for each class. Class 1 has the highest average values for real GDP per capita in 1960 and in 1995, as well as schooling, and the lowest for population growth, and the second highest saving rate, after class 2. Class 4 has the lowest real GDP per capita both in 1960 and in 1995, as well as the saving rate in terms of both physical and human capital on average. The highest population growth on average is in Class 3.

There is a large discrepancy among the classes in terms of average per capita income in both 1960 and 1995. Income per capita increased during the period on average for Classes 1, 2 and 3, however it decreased for Class 4 in real terms. We also observe an increase in the cross-section variation of income per capita within groups. On the other hand, the average group saving rates do not show as large a discrepancy as does average per capita incomes. Secondary school enrollment differs widely on average across groups. Population growth in average is low for those classes with high average per capita income. In Class 2, Saudi Arabia has a very high population growth rate relative to the rest of the group, once it is excluded, the average growth rate of population drops to 3.11%.

A point of caution is to be made. As indicated earlier, the results of a classification analysis depends crucially on the data set. Although the classification procedure used

in this paper is also prone to this issue, as the outcome is totally probabilistic, no deterministic statements are made. It is possible to strengthen the classification outcomes by using a concept similar to the one used in Hobijn and Franses (2000). In their paper, Hobijn and Franses (2000) use a concept that they call *cluster correlation*. They define cluster correlation as the degree of overlap of two outcomes for the same set of countries. Such an exercise is not carried out in this paper as it might be very difficult to obtain cluster correlation results for all classification models $\mathcal{M} \in S$. Therefore, posterior probabilities of the best alternative models are used as the sole basis of statistical comparison.[13]

4.3 Empirical Solow Model Within Groups

The classification summarized in Tables 6 and 7 displays the aspects of heterogeneity in the data. Rather than dividing the sample according to non-oil producers, intermediate group, and OECD countries as in MRW, the subsamples are formed according to the statistical distribution each country belongs to. This partitioning of the data set is based on the idea that the data is sampled from a heterogeneous population, and enables us to form systematic groupings accordingly. Within each subsample, countries have the same growth process, and thus, a separate regression analysis is performed on each subsample. Equations (10) through (13) are estimated for each subsample. The results are summarized below.

Table 8 reports the estimation results for equations (10) and (11). Note that the signs of the estimated coefficients do not turn out to be what the Solow model predicts once heterogeneity in the sample is accounted for. The estimates of equation (10) show a

[13] At this point, it might be useful to look at the Herfindahl index measure for the classification outcome of this section. The data set has 105 countries, and as the result of the Bayesian classification analysis carried out, the countries are partitioned into four groups, one with 24 countries, the second with 22, the third with 43 and the fourth with 16 countries indicating a Herfindahl index value of 0.71. This shows that the degree of diversity within the data set is quite high. (Note that this index is calculated using the formula $1-\sum_{i=1}^{n} s_i^2$ where n is the number of groups and s_i is the size of group i in proportion to the total sample size.) As stated above for the case of cluster correlation, the calculation of the Herfindahl index for all possible classification models $\mathcal{M} \in S$ might be quite difficult and is thus left out. The comparison of alternative classification models is based on the posterior probabilities of the models.

positive relationship between per capita income and population growth rate for Classes 1 and 2, and a negative relationship between the saving rate and per capita income for Class 2, though these estimates are not significant at conventional levels. Further, the estimated coefficients of saving rate for Class 1 and population growth for Class 4 are insignificant as well.

Once human capital is taken into account, population growth and per capita income is estimated to have a positive relationship for all of the subsamples as well as the whole sample. Note, however, that most of these estimates are not significant. Furthermore, there is a negative relationship between the saving rate and per capita income, this time for Classes 3 and 4. In fact, this is not inconsistent with the descriptive statistics in Table 7. It has been noted above that although there is a large discrepancy among the groups in terms of average income per capita, the saving rates do not seem to differ as much on average. This might suggest that the saving rates do not matter for income per capita in poorer countries. The coefficient estimates for schooling, on the other hand, have the expected signs. Classes 1 and 2, those that have the highest education levels on average, also have significant and positive estimates for schooling. The implication of these results is that the relationships predicted by the Solow model do not hold once the data set is partitioned into subsamples, each of which is generated by a different process.

Table 9 reports the estimation results for equations (12) and (13). The results indicate unconditional convergence for Class 1, Class 2, and Class 4 at 1.8%, 2%, and 1% respectively. Class 3, the largest, show divergence at a rate of -0.2% while there is divergence in the whole sample of -0.4%. Note however that these results are not significant for Class 3 and Class 4. Thus, there is not much evidence of unconditional convergence except for the groups that are toward the high-end of the distribution.

Equation (13) is estimated twice, once without schooling, and once with schooling. There is evidence of convergence for all subsamples as well as the whole sample in the first case, though it is not significant. The rates of convergence are 1%, 0.9%, 0.4%, 1%, and 0.1% for Class 1, 2, 3, 4, and the whole sample respectively. There is insignificant

evidence within-cluster convergence for Classes 2, 3, and 4, but significant for Class 1. Note, however, that in addition to the initial income per capita, for Class 2, the sign of the saving rate is also the opposite of what the Solow model predicts. Further, population growth is insignificant for Classes 1 and 4.

When schooling is included, again positive convergence within clusters and for the whole sample is found. The rates are 1.9%, 1.1%, 0.4%, 0.1% and 0.7% for Class 1, 2, 3, 4, and the whole sample respectively. These coefficients are again insignificant for Classes 2, 3, and 4. Contrary to the predictions of the Solow model, the regression outcomes imply a positive relation between population growth and growth of per capita income for Classes 1 and 2, and a negative relation between schooling and growth of per capita income for Class 4. Schooling is statistically insignificant in Class 3.

The results of this section indicate that once the heterogeneity in the sample is accounted for according to the underlying statistical distributions, the regression outcomes differ from what the Solow model predicts in terms of the signs of some of the estimated coefficients. In addition, it is possible to find divergence within some groups unconditionally. Estimated conditional convergence coefficients are larger and the results are more supportive of the Solow model when the partitioning is ad hoc.

5 Conclusion

The purpose of this study is to reconsider the empirical growth models by exploring the patterns of heterogeneity in the data set that would lead to the use of different regression models in different subsamples. The paper proposes the use of Bayesian classification to systematically reveal the patterns of heterogeneity in the data. The second step, then, is to use this information on heterogeneity in the cross-country growth regressions and in the tests of convergence hypothesis by performing a separate analysis for each subsample.

The data set used to illustrate the methodology includes the standard Solow growth model variables, which are also used by Mankiw *et al.* (1992). The results of the analysis

indicate that once the heterogeneity in the sample is accounted for according to the underlying statistical distributions, the regression outcomes differ from what the Solow model predicts in terms of the signs of some of the estimated coefficients.

The important point to note is that this method does not explain the underlying reasons for the differences between the groups. It is based on the idea that the data comes from different statistical distributions, however it cannot characterize the determinants of the differences. In addition, the analyses indicate that although countries in the same group experience the same growth process, there is no significant evidence supporting within group convergence in income levels, except perhaps for the group that includes all the advanced economies. The reasons for these two issues are beyond the scope of this paper, and are left as future research questions to be explored.

The findings imply that it would be misleading to run cross-country growth regressions on the whole sample without taking the subsample differences into account. In addition, once the underlying factors of the differences across groups are identified, more accurate policy prescriptions can be made to improve the conditions and to enhance growth in the less developed countries so that they catch up with the rich.

References

Ardıç, O. P. (2004). The dynamics of the distribution of standard of living across countries. Mimeo, Boğaziçi University, Department of Economics.

Barro, R. (1991). Economic growth in a cross-section of countries. *Quarterly Journal of Economics*, **106**, 407–443.

Barro, R. (1996). Democracy and growth. *Journal of Economic Growth*, **1**, 1–27.

Barro, R. and Lee, J. (1993). International comparisons of educational attainment. *Journal of Monetary Economics*, **32**, 363–394.

Barro, R. and Sala-i Martin, X. (1995). *Economic Growth*. McGraw-Hill.

Bernard, A. and Durlauf, S. (1996). Interpreting the tests of the convergence hypothesis. *Journal of Econometrics*, **71**, 161–172.

Brock, W. and Durlauf, S. (2001). Growth empirics and reality. *The World Bank Economic Review*, **15**, 229–272.

Canova, F. (1999). Testing for convergence clubs in income per capita: A predictive density approach. *Mimeo, University of Pompeu Fabra*.

Dillon, W. R. and Goldstein, M. (1984). *Multivariate Analysis*. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc.

Durlauf, S. (2000). Econometric analysis and the study of economic growth: A skeptical perspective. *Macroeconomics and the Real World,* edited by R. Backhouse and A. Salanti. Oxford: Oxford University Press.

Durlauf, S. and Johnson, P. (1995). Multiple regimes and cross-country growth behavior. *Journal of Applied Econometrics*, **10**, 365–384.

Durlauf, S. and Quah, D. (1999). The new empirics of economic growth. *Handbook of Macroeconomics,* North-Holland.

Durlauf, S., Kourtelos, A., and Minkin, A. (2001). The local Solow growth model. *European Economic Review*, **45**, 928–940.

Hanson, R., Stutz, J., and Cheeseman, P. (1991). Bayesian classification theory. *Technical Report FIA-90-12-7-01, NASA Ames Research Center*.

Hobijn, B. and Franses, P. H. (2000). Asymptotically perfect and relative convergence of productivity. *Journal of Applied Econometrics*, **15**, 59–81.

Hobijn, B. and Franses, P. H. (2001). Are living standards converging? *Structural Change and Economic Dynamics*, **12**, 171–200.

Kourtellos, A. (2001). A projection pursuit approach to cross country growth data. *Mimeo, University of Wisconsin-Madison, Department of Economics*.

Levine, R. and Renelt, D. (1992). A sensitivity analysis of cross-country growth regressions. *American Economic Review*, **82**, 942–963.

Mankiw, G., Romer, D., and Weil, D. (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics*, **107**, 407–437.

Mauro, P. (1995). Corruption and growth. *Quarterly Journal of Economics*, **110**, 681–713.

Quah, D. (1993). Galton's Fallacy and the tests of the convergence hypothesis. *Scandinavian Journal of Economics*, **94**, 427–443.

Quah, D. (1996a). Convergence empirics across economies wit (some) capital mobility. *Journal of Economic Growth*, **1**, 95–124.

Quah, D. (1996b). Empirics for economic growth and convergence. *European Economic Review*, **40**, 1353–1375.

Quah, D. (1997). Empirics for growth and distribution: Stratification, polarization, and convergence clubs. *Journal Economic Growth*, **2**, 27–59.

Solow, R. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, **70**, 65–94.

Stutz, J. and Cheeseman, P. (1996). Autoclass - a Bayesian approach to classification. *Maximum Entropy and Bayesian Methods* edited by J. Skilling and S. Sibisi, Kluwer Academic Publishers.

Sala-i Martin, X. (1997). I just ran two million regressions. *American Economic Review, Papers and Proceedings*, **87**, 178–183.

	y_{1960}	y_{1995}	s_h (%)	s_k (%)	n (%)
Mean	2,251	5,038	45.78	21.32	3.37
Std.Dev.	2,154	5,194	29.32	5.81	0.05
Minimum	257	225	3.69	7.45	0.07
Maximum	9,895	18,975	104.44	37.33	10.45
Range	9,638	18,750	100.75	29.88	10.38
# of Obs.	105				

Table 1: Desctiptive Statistics.

Dependent Variable: log GDP/capita in 1995

Sample	Non-Oil	Intermediate	OECD
# of Obs.	100	71	18
Constant	3.984	3.848	5.900
	(1.182)	(1.270)	(2.186)
$\ln(I/GDP)$	1.652	1.529	0.101
	(0.286)	(0.379)	(0.537)
$\ln(n+g+\delta)$	-2.609	-2.648	-1.315
	(0.379)	(0.388)	(0.689)
$\overline{R}^2$	0.537	0.534	0.097

Sample	Non-Oil	Intermediate	OECD
# of Obs.	100	71	18
Constant	6.846	8.408	11.037
	(0.910)	(1.089)	(1.760)
$\ln(I/GDP)$	-1.016	-0.418	0.558
	(0.323)	(0.400)	(0.588)
$\ln(n+g+\delta)$	0.302	0.134	-0.143
	(0.253)	(0.327)	(0.346)
$\ln(school)$	0.940	1.249	1.599
	(0.100)	(0.159)	(0.334)
$\overline{R}^2$	0.756	0.756	0.633

Table 2: The OLS estimates of equations (10) and (11)
Note: Standard errors are in parantheses.

Dependent Variable: log difference of GDP/capita 1960-95			
Unconditional Convergence			
Sample	Non-Oil	Intermediate	OECD
# of Obs.	100	71	18
Constant	-0.653	0.397	4.029
	(0.505)	(0.597)	(0.882)
$\ln(y60)$	0.169	0.042	-0.365
	(0.068)	(0.079)	(0.103)
$\overline{R}^2$	0.049	-0.010	0.404
Implied λ	-0.004	-0.001	0.013
	(0.002)	(0.002)	(0.005)

Table 3: The OLS estimates of equation (12)
Note: Standard errors are in parantheses.

Dependent Variable: log difference of GDP/capita 1960-95

Conditional Convergence			
Sample	Non-Oil	Intermediate	OECD
# of Obs.	100	71	18
Constant	1.639	2.372	3.278
	(0.655)	(0.721)	(0.963)
$\ln(I/GDP)$	1.332	1.542	0.803
	(0.155)	(0.212)	(0.239)
$\ln(n+g+\delta)$	-0.548	-0.672	-0.551
	(0.244)	(0.271)	(0.301)
$\ln(y60)$	-0.042	-0.131	-0.319
	(0.062)	(0.071)	(0.080)
$\overline{R}^2$	0.495	0.489	0.702
Implied λ	0.001	0.004	0.011
	(0.002)	(0.002)	(0.003)
Sample	Non-Oil	Intermediate	OECD
# of Obs.	100	71	18
Constant	3.152	4.238	5.167
	(0.728)	(0.973)	(1.689)
$\ln(I/GDP)$	0.895	1.057	0.609
	(0.185)	(0.270)	(0.273)
$\ln(school)$	0.348	0.432	0.441
	(0.091)	(0.159)	(0.328)
$\ln(n+g+\delta)$	-0.359	-0.287	-0.173
	(0.233)	(0.395)	(0.406)
$\ln(y60)$	-0.228	-0.301	-0.442
	(0.076)	(0.092)	(0.120)
$\overline{R}^2$	0.558	0.533	0.718
Implied λ	0.007	0.010	0.017
	(0.003)	(0.004)	(0.006)

Table 4: The OLS estimates of equation (13)
Note: Standard errors are in parantheses.

# of Classes	1	2	3	4	5	6
log probability	-4063	-4012	-4022	-4002	-4012	-4020

Table 5: Alternative Classification Models

Class 1	Class 2	Class 3		Class 4
Australia	Argentina	Algeria	Indonesia	Burkina Faso
Austria	Botswana	Bangladesh	Iran	Burundi
Bahamas	Chile	Belize	Kenya	C. Afr. Rep.
Barbados	China	Benin	Mexico	Chad
Belgium	Guyana	Bolivia	Morocco	Congo, D.R.
Canada	Hong Kong	Brazil	Nepal	Guinea-Bissau
Denmark	Jamaica	Cameroon	Nicaragua	Haiti
Finland	Korea, Rep.	Colombia	Nigeria	Madagascar
France	Lesotho	Congo, Rep.	Pakistan	Malawi
Greece	Malaysia	Costa Rica	Panama	Mali
Hungary	Mauritius	Cote d'Ivoire	Paraguay	Mauritania
Iceland	Oman	Dominican R.	Peru	Mozambique
Israel	Romania	Ecuador	Philippines	Niger
Italy	Saudi Arabia	Egypt	Senegal	P.N. Guinea
Japan	Singapore	El Salvador	S.Africa	Rwanda
Malta	Sri Lanka	Ethiopia	Sudan	Zambia
Netherlands	Suriname	Fiji	Swaziland	
New Zealand	Syria	Gambia, the	Togo	
Norway	Thailand	Ghana	Turkey	
Spain	Trinidad & Tobago	Guatemala	Uganda	
Sweden	Tunisia	Honduras	Zimbabwe	
UK	Venezuela	India		
USA				
Uruguay				

Table 6: Classification Results: List of Countries

	Class 1	Class 2	Class 3	Class 4
GDP/Capita-60 - Average	5,234	2,059	1,247	736
GDP/Capita-60 - Standard Deviation	2,212	1,665	621	280
GDP/Capita-60 - Minimum	1,374	313	257	380
GDP/Capita-60 - Maximum	9,895	6,338	2,946	1,235
GDP/Capita-95 - Average	12,498	5,763	2,143	630
GDP/Capita-95 - Standard Deviation	3,728	4,350	1,277	346
GDP/Capita-95 - Minimum	4,870	1,142	331	225
GDP/Capita-95 - Maximum	18,975	18,051	5,919	1,787
Schooling - Average	87.98	50.51	33.28	9.56
Schooling - Standard Deviation	9.77	16.65	15.46	5.34
Schooling - Minimum	69.02	20.22	8.59	3.69
Schooling - Maximum	104.44	81.12	64.84	19.86
Saving Rate - Average	22.95	26.49	19.75	15.95
Saving Rate - Standard Deviation	4.07	4.73	4.60	6.04
Saving Rate - Minimum	15.97	19.65	11.25	7.45
Saving Rate - Maximum	32.46	37.33	33.70	26.67
Population Growth - Average	1.10	3.44	4.40	3.88
Population Growth - Standard Deviation	0.03	0.07	0.03	0.03
Population Growth - Minimum	0.07	0.66	2.90	2.52
Population Growth - Maximum	4.64	10.45	7.37	5.78

Table 7: Descriptive Statistics for the Classes

Dependent Variable: log GDP/capita in 1995

Sample	Class 1	Class 2	Class 3	Class 4	All Sample
# of Obs.	24	22	43	16	105
Constant	10.565	8.891	3.083	4.075	5.261
	(1.605)	(2.415)	(1.929)	(2.124)	(1.159)
$\ln(I/GDP)$	0.044	-0.864	1.461	0.653	1.674
	(0.447)	(0.955)	(0.321)	(0.243)	(0.243)
$\ln(n+g+\delta)$	0.399	0.657	-2.874	-1.447	-2.126
	(0.483)	(0.665)	(0.767)	(0.857)	(0.373)
$\overline{R}^2$	-0.026	0.030	0.556	0.344	0.454
Sample	Class 1	Class 2	Class 3	Class 4	All Sample
# of Obs.	24	22	43	16	105
Constant	11.306	14.803	5.341	4.003	8.001
	(1.086)	(3.230)	(0.703)	(2.207)	(0.854)
$\ln(I/GDP)$	0.525	2.081	-1.472	-1.535	-0.563
	(0.325)	(0.829)	(0.260)	(0.919)	(0.302)
$\ln(n+g+\delta)$	0.090	0.131	0.270	0.640	0.243
	(0.300)	(0.941)	(0.253)	(0.254)	(0.249)
$\ln(school)$	2.360	1.363	0.736	0.066	1.025
	(0.457)	(0.556)	(0.318)	(0.179)	(0.099)
$\overline{R}^2$	0.724	0.233	0.643	0.298	0.732

Table 8: The OLS estimates of equations (10) and (11) for each subsample
Note: Standard errors are in parantheses.

Dependent Variable: log difference of GDP/capita 1960-95

Unconditional Convergence					
Sample	Class 1	Class 2	Class 3	Class 4	All Sample
# of Obs.	24	22	43	16	105
Constant	4.926	4.731	0.092	1.809	-0.394
	(0.849)	(1.258)	(0.688)	(1.664)	(0.502)
ln($y60$)	-0.473	-0.499	0.056	-0.306	0.136
	(0.100)	(0.171)	(0.098)	(0.254)	(0.068)
$\overline{R}^2$	0.281	0.264	-0.012	0.029	0.028
Implied λ	0.018	0.020	-0.002	0.010	-0.004
	(0.005)	(0.010)	(0.003)	(0.010)	(0.002)
Conditional Convergence					
Sample	Class 1	Class 2	Class 3	Class 4	All Sample
# of Obs.	24	22	43	16	105
Constant	4.960	8.001	-2.436	0.952	2.095
	(1.163)	(2.011)	(1.086)	(1.591)	(1.591)
ln(I/GDP)	0.990	1.588	0.652	0.719	1.364
	(0.274)	(1.106)	(0.176)	(0.161)	(0.157)
ln($n+g+\delta$)	-0.002	1.088	-2.095	-0.881	-0.331
	(0.266)	(0.565)	(0.389)	(0.584)	(0.225)
ln($y60$)	-0.304	-0.282	-0.139	-0.293	-0.021
	(0.096)	(0.228)	(0.078)	(0.168)	(0.060)
$\overline{R}^2$	0.541	0.339	0.430	0.601	0.462
Implied λ	0.010	0.009	0.004	0.010	0.001
	(0.004)	(0.009)	(0.003)	(0.007)	(0.002)
Sample	Class 1	Class 2	Class 3	Class 4	All Sample
# of Obs.	24	22	43	16	105
Constant	6.703	13.471	-2.435	0.153	3.738
	(1.400)	(2.559)	(1.348)	(1.313)	(0.730)
ln(I/GDP)	0.766	2.360	0.651	0.802	0.910
	(0.281)	(0.974)	(0.265)	(0.134)	(0.187)
ln($school$)	0.962	1.249	0.000	-0.308	0.371
	(0.487)	(0.437)	(0.150)	(0.112)	(0.094)
ln($n+g+\delta$)	0.152	2.369	-2.094	-0.266	-0.135
	(0.261)	(0.655)	(0.463)	(0.520)	(0.216)
ln($y60$)	-0.483	-0.323	-0.139	-0.035	-0.223
	(0.128)	(0.193)	(0.093)	(0.165)	(0.076)
$\overline{R}^2$	0.687	0.527	0.418	0.741	0.530
Implied λ	0.019	0.011	0.004	0.001	0.007
	(0.007)	(0.008)	(0.003)	(0.005)	(0.003)

Table 9: The OLS estimates of equations (12) and (13) for each subsample
Note: Standard errors are in parantheses.