NO ONE TRUE PATH: UNCOVERING THE INTERPLAY BETWEEN GEOGRAPHY, INSTITUTIONS, AND FRACTIONALIZATION IN ECONOMIC DEVELOPMENT [*]

Chih Ming Tan

This Draft: February 8, 2008

Abstract

Do institutions rule when explaining cross-country divergence? By employing regression tree analysis to uncover the existence and nature of multiple development clubs and growth regimes, this paper finds that to a large extent they do. However, the role of ethnic fractionalization cannot be dismissed. The findings suggest that sufficiently high-quality institutions may be necessary for the negative impact on development from high levels of ethnic fractionalization to be mitigated. Interestingly, I find no role for geographic factors; neither those associated with climate nor physical isolation, in explaining divergence. There is also no evidence to suggest a role for religious fractionalization.

Keywords: Regression trees, threshold regression, economic growth, institutions, geography, fractionalization

JEL Classifications: C45, C49, C52, N15, N16, N17, O10

---

[*] Department of Economics, Braker Hall, Tufts University, 8 Upper Campus Road, Medford, MA 02155. Email: chihming.tan@tufts.edu. I thank Buz Brock, Steven Durlauf, Dan Henderson, Yannis Ioannides, Paul Johnson, Louise Keely, Andros Kourtellos, Chris Lintz, Rody Manuelli and, Chris Papageorgiou for insightful comments and discussions. Wei-Yin Loh offered invaluable advice on the implementation of the regression tree methods used in this paper. I am also deeply grateful to the MacArthur Research Network on Social Interactions and Economic Inequality for generous research support under the Program of Fellowships for Junior Scholars.

1. Introduction

The recent empirical growth literature has seen serious attempts at evaluating the question of whether certain non-traditional growth determinants may be responsible for the observed divergence in the cross-country distribution of income (see, Quah, 1997, and Canova, 2004). One characteristic of this literature has been the recasting of the growth process as a system that exhibits a "hierarchy of timescales"[1] whereby growth is determined by slow- and fast-moving variables.

In this view, the familiar neoclassical growth determinants (see, Solow, 1956, and Mankiw et al., 1992) – physical and human capital accumulation rates, and population growth rates – constitute the "proximate" determinants of growth while slow-moving variables such as a country's geography (e.g., Bloom and Sachs, 1998, Gallup et al. 1999), the quality of its institutions (e.g., Hall and Jones, 1999, Acemoglu et al. 2001), and the degree of fractionalization (e.g., Easterly and Levine, 1997, Alesina et al., 2003) in its society are considered to be "fundamental" determinants.

The recent literature has seen much controversy over the relative empirical salience of these fundamental determinants to growth. Efforts to assess the evidence in favor of each of these fundamental theories have suffered from three related and important shortcomings. First, most studies argue from a priori positions that assume mono-causality. For example, in the highly contentious "institutions versus geography" debate (see, Rodrik et al., 2004, Easterly and Levine, 2003, Sachs, 2003) the aim is to test if institutions "rule" and, therefore, whether geography's importance is limited solely to its historical role in determining the initial quality of institutions that a country inherits at birth. Second, virtually all of the research assumes away substantial heterogeneity (e.g., interactions between theories) and nonlinearity across observational units. Attempts to resolve the institutions versus geography debate, for instance, have centered largely upon efforts to uncover partial correlations between various proxies for fundamental determinants with per capita income or

[1] Terminology due to Brock (2001).

conditional growth rates within a linear framework. Little attention has been paid in the literature to the question of whether the linear model adequately captures structure in the data. Finally, when researchers have paid attention to heterogeneity and nonlinearity, they have typically limited their investigation to a small number of alternative model specifications before settling on a particular specification that is then reported. Barro (1996), for instance, focuses on the nonlinear effect of democracy on growth, but excludes other possible nonlinearities and interaction effects.

The combined outcome of the above strategy is to imply strong prior knowledge on the part of researchers about the correct (econometric) model for growth. Instead of taking into account the very large set of possible models that could be generated by accounting for possible nonlinearities and interactions between covariates, researchers essentially narrow their focus, a priori, to a very small number of alternative models. However, as pointed out by Brock and Durlauf (2001), growth theories are fundamentally "open-ended". By open-ended, Brock and Durlauf mean that the fact that one theory (e.g., institutions) may be salient to growth does not automatically exclude some other theory or theories (e.g., geography or ethnic fractionalization) from also being important. It also does not exclude the possibility that their interaction may be important. The deep insight by Brock and Durlauf is that theory open-endedness implies uncertainty over the correct model specification for growth, and therefore, any assessment of the marginal impact of growth theories on economic performance needs to place the full universe of possible alternative models under consideration.

The main contribution of this paper is to address the question of model uncertainty within the nonlinear framework[2]. In this paper, I employ sample splitting and threshold estimation methods (notably, regression trees and Hansen's threshold regression) to identify important fundamental

[2] In other work, for instance, Durlauf, Kourtellos, and Tan (2007), we have considered the problem of model uncertainty in the linear framework.

growth determinants as well as their salient interactions. Many growth theories suggest that the growth process may be characterized by threshold nonlinearities (see, for instance, Galor and Moav, 2002, Galor, 2005, Chamon and Kremer, 2006). Sample splitting methods iteratively split the sample of observations into increasingly homogeneous subsets. At each stage, a threshold variable is chosen from the set of fundamental determinant proxies and a threshold value is determined to facilitate the splitting of the sample. The result of the iterative sample splitting procedure is to deliver groups of countries whose members share meaningful similarities in the way fundamental determinants influence economic outcomes without the need for us to impose any a priori structure on the number of these groups. The results are also structurally interpretable in the sense that they reveal the relative importance of particular fundamental determinants to countries in each of these groups.

The findings in the paper provide robust support for the view that institutions and their interaction with ethnic fractionalization are central to explaining cross-country economic divergence. Higher levels of ethnic fractionalization have no impact on development for the group of countries with high quality institutions. For countries with quality of institutions below a threshold level, however, greater ethnic fractionalization is associated with substantially lower levels of development. Sufficiently high-quality institutions are necessary to mitigate the negative impact on development from high levels of ethnic fractionalization. Interestingly, I find no role for geographic factors; neither those associated with climate nor geographic isolation, in explaining divergence. There is also no evidence to suggest an important role for religious fractionalization. Finally, the findings in this paper affirm earlier work (e.g., Masanjala and Papageorgiou, 2007) in the literature that sets apart the development process of Sub-Saharan Africa from the rest of the world.

The remainder of the paper is organized as follows. Section 2 of this paper provides an empirical framework for the discussion. I model the timescale effect of fundamental determinants on economic performance as a search for the existence of multiple development clubs and growth regimes. Section 3 describes the regression tree algorithm employed in this paper; i.e., Generalized Unbiased Interaction Detection and Estimation (GUIDE). Sections 4 and 5 discuss the data and estimation results respectively. Finally, Section 6 concludes.

2. Development Clubs and Growth Regimes

I model the influence of fundamental determinants on economic outcomes in two ways.

2.1 *Development Clubs*

I first consider the case where the dependent variable is the period ending level of per capita income (as opposed to the growth rate, which I consider in the next subsection). Hall and Jones (1999) have advocated the use of such levels regressions on the grounds that differences in long-run levels of income are the most relevant measures of welfare and suggest that the low correlation of growth rates across decades may imply that observed variations in cross-country growth rates may be mostly transitory. This levels regression approach also characterizes much of the literature in the "institutions versus geography" debate.

Our aim is to classify the long-run development levels of countries into groups defined by threshold values for fundamental determinants[3]. Formally, suppose there are b mutually exclusive partitions in the space of fundamental determinants, $\{A_j\}_{j=1}^{b}$. That is, $\forall j \neq l$,

[3] Bloom, Canning, and Sevilla (2003), which employs Gaussian mixture models, is a closely related work. However, Bloom et al only consider geography (latitude) as an indexing variable for subgroups of countries. They are therefore not able to distinguish the relative importance to development of the three classes of fundamental determinants. This paper

$$A_j \cap A_l = \varnothing \tag{1}$$

and,

$$\bigcup_{j=1}^{b} A_j = Z \tag{2}$$

where Z is the space of fundamental determinants. As a simple example to illustrate how we would interpret thresholds in the space of fundamental determinants, let us suppose that fundamental determinants took on discrete values. So, suppose $z = (Instit, Frac, Geog)$ where $Instit$ takes on values $\{HI, LI\}$, $Frac$ takes on values $\{HF, LF\}$, and $Geog$ takes on values $\{Tr, Tm\}$. A possible set of partitions is $\{(Tr, Tm, LI, HF), (Tr, Tm, LI, LF), (Tr, Tm, HI, HF, LF)\}$. That is, there is a "low quality institutions-high fractionalization" partition, a "low quality institutions-low fractionalization" partition, and a "high quality institutions" partition. Geography, in this example, is not salient in partitioning countries into groups. This example makes clear how partitioning the space of threshold variables potentially reveals both the importance of particular threshold variables as well as the role of interactions between threshold variables in characterizing heterogeneity.

Since we are only concerned with explaining levels of long-run development, the regression specification within each partition of the fundamental determinants space, A_j, for $j = 1, \ldots, b$, is therefore simply a (piece-wise) constant model corresponding to,

$$E(y_i \mid z_i \in A_j) = y_j^*, \qquad j = 1, \ldots, b \tag{3}$$

where y_i is real per capita income for country i, and $z_i \in Z$ is the vector of values for fundamental determinants for country i. The number of development clubs in this case is b. One of the tasks of estimation will be to uncover the exact value of b.

---

extends their contribution by considering all three fundamental determinants; geography, institutions, and fractionalization.

2.2 *Multiple Growth Regimes*

Next, I explicitly model the effects of fundamental determinants in terms of their role in driving heterogeneity in production technologies. As opposed to the levels regression above, growth regressions are interpretable within the context of the (augmented) neoclassical growth model. In the context of this paper, growth regressions also allow us to relate our findings to those in the large literature on multiple growth regimes (e.g., Canova, 2004, Desdoigts, 1999, Durlauf and Johnson, 1995, and Durlauf, Kourtellos, and Minkin, 2001).

We start formally with a model of the neoclassical production function for each country i given by,

$$y_i = f_\varepsilon^j(k_i) \quad \textit{iff} \quad z_i \in A_j,\ j = 1,\ldots,b \tag{4}$$

where y is real per capita income, f_ε is a neoclassical production function, k is a vector of per capita capital stocks that all depreciate at a common rate, δ, and $z \in Z$ is the vector of fundamental determinants. If we assume further that the production function is Cobb-Douglas, then, following Mankiw, Romer, and Weil (1992), growth around each steady state would be given to a first order approximation by a growth regime such as

$$g_i = \alpha_j + \beta_j^k \cdot \ln(s_i^k) + \beta_j^h \cdot \ln(s_i^h) + \beta_j^n \cdot \ln(n_i + \delta + \xi) + \beta_j^0 \cdot \ln(y_i^0) + \varepsilon_i \tag{5}$$

iff $z_i \in A_j$ for $j = 1,\ldots,b$.

Here, g_i is the difference in log per capita real GDP between the start and end years of a given time interval for country i, s_i^k and s_i^h are, respectively, the average rates of physical and human capital accumulation across the time interval, $(n_i + \delta + \xi)$ is the sum of, respectively, the average population growth rate across the time interval, the depreciation rate for physical and human capital, and the rate of exogenous technological growth, y_i^0 is per capita real GDP in the initial year, and ε is a mean 0, variance σ_j^2 innovation. The coefficient on log initial per capita GDP, β_j^0, if negative, is typically interpreted as an indication of how quickly poorer countries are catching up with richer countries within the group.

3. Regression Tree Analysis and Generalized Unbiased Interaction Detection and Estimation (GUIDE)

Our task is therefore to estimate the number of, respectively, development clubs or growth regimes, b, the threshold values and fundamental determinant choices that define the partitions, $\left\{A_j\right\}_{j=1}^{b}$, and the set of regression parameters associated with development clubs, $\left\{y_j^*\right\}_{j=1}^{b}$, or growth regimes, Θ; where Θ is defined as the set $\left\{\alpha_j, \beta_j^k, \beta_j^h, \beta_j^n, \beta_j^0, \sigma_j^2\right\}_{j=1}^{b}$.

Regression tree analysis provides a computationally efficient method for fulfilling this purpose[4]. Regression tree methods are standard in the statistical learning literature and bear deep similarities to recent sample splitting and threshold estimation methods in the econometrics literature (see, in particular, Hansen, 2000, as well as Gonzalo and Pitarakis, 2002). Regression tree methods are appropriate tools for the job because they routinely handle the case of multiple threshold variables as well as multiple thresholds. They have also been shown to be consistent in the sense that as the number of observations gets large, regression trees reproduce the "true" set of sample splits (see Breiman, Friedman, Olsen, and Stone, 1984). Their weakness, however, lies in the

[4] In terms of methodology, my approach is closest to Durlauf and Johnson (1995) who also use regression tree techniques in their analysis.

lack of available asymptotic results that would be useful for conducting inference on threshold variable choices and threshold value estimates. The econometrics literature has sought to correct this within the context of test-based sequential sample splitting models (see Hansen, 1999, Hansen, 2000). However, results such as confidence intervals derived in these settings are generally restricted to the single threshold variable-single threshold case. Nevertheless, our regression tree results (see Section 5) confirm that these inferential results are applicable in the multiple regimes context. I therefore also report results for confidence intervals for threshold estimates using Hansen's methodology in Section 5 below.

The specific regression tree algorithm I employ to uncover development clubs and growth regimes is GUIDE. Loh (2002) is the key reference. GUIDE is an extension of the Classification and Regression Tree (CART) methodology by Breiman et al. (1984). GUIDE's innovation is to minimize potential biases in variable selection and interaction detection in CART. These biases arise because the "greedy" search algorithm employed by CART for tree splitting selects threshold variables with larger support points more frequently regardless of fit (see Doyle, 1973). GUIDE overcomes the bias problem by replacing the "greedy" search algorithm with a simple LM test of linear fit.

GUIDE builds a tree by employing the following tree building algorithm: First start with the set of all observations. To locate a threshold variable, GUIDE first fits the regression equation (e.g. constant or linear model) to the data and obtains the regression residuals. GUIDE then creates a contingency table for each candidate threshold variable (each fundamental determinant variable in this case) with the latter's support partitioned into quartiles as the columns of the contingency table, and negative or positive regression residuals as the rows. Each cell of the contingency table contains a count for the number of regression residuals that are either positive or negative for observations that correspond to each quartile as given by the columns of the contingency table. A Chi-square test for linear fit is then carried out to determine if there is evidence of curvature. This is done for each candidate threshold variable. The p-values are obtained using bootstrap. The candidate threshold variable corresponding to the lowest p-value is selected to split the data for this stage of the tree. Once a threshold variable has been selected, a threshold value is arrived at by finding a value for the threshold variable that minimizes the joint classical linear regression sum of squared errors across the resulting two subgroups. Parameters for the regression equation are estimated by concentration.

GUIDE then applies the above threshold selection procedure iteratively to each newly formed subgroup. At each stage, the given subset of observations is further divided into two subgroups. The procedure stops only after a pre-set value for the minimum number of observations in a subgroup has been breached. The result is the construction of an "overly large" tree. To deal with the problem of over-fitting, the overly large tree is then pruned back using a criterion, similar to a generalized information criterion, that maximizes overall fit while penalizing for complexity (i.e., the number of subgroups) to arrive at the final tree.

Figure 1, due to Hastie, Tibshirani, and Friedman (2001), provides a simple schematic to show what a regression tree algorithm such as GUIDE accomplishes. In experiments, Loh (2002) shows that there is negligible bias in the selection of threshold variables. This provides some confidence in the interpretability of the uncovered structure in small samples. Loh also shows that GUIDE provides good out-of-sample prediction performance when compared to other machine learning algorithms across a wide range of data sets. In particular, with bootstrap aggregation, GUIDE provides the best predictive performance out of the set of learning algorithms that delivers interpretable structure (including multivariate adaptive regression splines (MARS)). The reader is referred to the Technical Appendix for details on the GUIDE procedure.

4. Data

The dependent variable for the development clubs regression is the log of per capita real purchasing power-adjusted GDP[5] for 1999 for each country. Since this is a piece-wise constant regression model, the independent variables are proxies for fundamental determinants (described below) that are used for tree-splitting.

For the growth regimes regression, we employ panel data for each country over two 20-year time intervals, 1960-79 and 1980-99. The regression model in this case is piece-wise linear. The dependent variable is the difference between the logs of real per capita GDP for the start and end years of each of these intervals. Three of the five covariates that enter the (piece-wise) regression equation linearly, the log net depreciation rate[6] (MNGD), log investment share (MINV), and log schooling (MSCH15) are logs of averages taken across each of the two time intervals. The fourth covariate, log initial per capita income (MGDP0), is the log of real per capita GDP for the initial year in each of the two time intervals. Finally, a dummy variable for the period 1960-79 (DUM6079) is also included. The schooling variable corresponds to the "average years of secondary and higher level schooling for males aged 15 and above" calculated using data from Barro and Lee (2000). All of the other national accounting data come from the recently released Penn World Table 6.1 (PWT

[5] GDP figures used in this paper correspond to RGDPL in PWT 6.1. RGDPL is obtained by adding up consumption, investment, government and exports, and subtracting imports in any given year. The given year components are obtained by extrapolating the 1996 values in international dollars from the Geary aggregation using national growth rates. It is a fixed base index where the reference year is 1996.

[6] The net depreciation rate refers to the sum of the population growth rate, the rate of depreciation of physical and human capital, and the exogenous rate of technological change. By convention, the sum of the latter two variables is set to 0.05. Data for population growth rates are from PWT 6.1.

6.1) dataset (see Heston, Summers, and Aten, 2002). We turn now to data for geography, institutions, and fractionalization.

Following Diamond's (1997) seminal work, some growth economists have recently argued that disadvantageous geography has long-term consequences for a country's development. In particular, the impact of climate on agricultural productivity and disease ecology, and consequently on health, educational achievement, productivity, and the nature of policy regimes[7], has taken on a prominent role in this new literature. To proxy for climate, I use data from Harvard University's Center for International Development (CID) on the percentage of a country's land area that is classified as a tropical eco-zone (ZTROPICS). Masters and McMillan (2001) have also argued strongly for an important role for the prevalence of ground frost during the winter season for soil renewal and fertility as well as in eradicating disease vectors. I include their most favored variable, the proportion of a country's land area that experiences more than 5 frost-days per month in winter (FROST5), as an alternative proxy for climate. Other proponents of geography have emphasized the role of geographic isolation in inhibiting development (see, Frankel and Romer, 1999, and Radelet and Sachs, 1998). I use data (also from the CID) measuring the percentage of a country's land area within 100 km of an ice-free coast (LCR100KM) to proxy for geographic isolation[8].

The case for the importance of economic institutions in affecting long-run growth and development outcomes arose to some extent as a response to this new "geographic determinism". A

---

[7] For example, Gallup, Sachs and Mellinger (1999) argue that it is precisely the poor natural endowments of countries in the tropics that lead "the [tax revenue maximizing] sovereign [to trade] quick gain for future loss". Masters and McMillan (2003) and McMillan (2001) put forth similar views in the context of Sub-Saharan Africa.

[8] Separating the two dimensions of geographic heterogeneity, climate and geographic isolation, in empirical exercises as opposed to applying a generalized geographic proxy for both, like latitude, is strongly advocated by the literature (see McArthur and Sachs, 2001).

large number of studies[9] have shown that the quality of institutions has potentially crucial consequences for investment and productivity. It has also been posited that volatile macroeconomic policies, resulting in uncertainty in the economic environment, is likely to be symptomatic[10] of poor underlying institutional quality (see, in particular, Acemoglu et al., 2003). The main argument here is that economic performance is a direct consequence of economic institutions, so all other factors are ultimately interesting only in the context of how they affect the evolution of such economic institutions. The prediction therefore is that economic performance will be conditionally independent of factors like geography once economic institutions are accounted for.

Two variables are considered as alternative measures of the quality of institutions. The first, ICRG8497, measures the average level of country risk across the years 1984 to 1997. This is a comprehensive measure of institutional quality that aggregates across five variables measuring the quality of the bureaucracy, corruption in government, rule of law, expropriation risk, and repudiation of contracts by government. The second, EXPROP8497, measures specifically the risk of expropriation. Both of these variables come from the IRIS-3 dataset by Knack and Keefer. The decision to use both variables derives from the desire to facilitate the comparability of results across work in the literature. The correlation between these variables is very high at over 0.8.

Finally, building on work by Persson and Tabellini (1994) and others, some researchers have attributed under-development to the multiplicity of population subgroups; defined by differences in socio-cultural factors such as racial features, language, and religion, within a country. Proponents of fractionalization do contend that greater polarization potentially leads to competitive rent seeking

[9] See, for instance, Acemoglu, Johnson and Robinson (2001), Hall and Jones (1999), La Porta, Lopez-de-Silanes, Shleifer, and Vishny (1999), and Knack and Keefer (1995).

[10] As Easterly and Levine (2003) put it, "Bad policies would be kind of like a high fever from a bacterial infection. Packing the patient in ice would bring down the fever but does not cure the infection".

activities by groups in power at the expense of society as a whole, leading thereby to a degradation of economic institutions. However, they also argue that fractionalization has important implications for development outside of its effects on the institutionally-driven incentives to produce or divert. High levels of fractionalization could lead to disagreements over the desirability of the type and level of public goods to be provided leading to a socially suboptimal allocation of growth-critical public goods such as public schooling and physical infrastructure. This could occur because some public goods are viewed as benefiting one group more than another. Alesina, Baqir, and Easterly (1999), for instance, find that public goods provision is negatively correlated with fractionalization using census data. Also, since production is typically a joint activity across social groups, the willingness to reach a consensus and to cooperate in production becomes a factor in determining production possibilities. However, there is evidence that trust does not translate easily across ethnic lines (see, for instance, Alesina and La Ferrara, 2002, and Glaeser, Laibson, Sheinkman, and Soutter, 2000).

Easterly and Levine (1997) introduced the ethno-linguistic fractionalization index, ELF60, as a measure of ethnic fractionalization. ELF60 measures the probability in 1960 that two randomly selected people from a given country will not belong to the same ethno-linguistic subgroup. Until recently, ELF60 featured prominently as the default proxy choice for fractionalization. In a recent paper, Alesina, et al. (2003) argue that linguistic diversity alone may not adequately capture the different aspects of fractionalization in society. In particular, they argue that the degree to which differences in racial features, language, or religion approximate such fractionalization may vary from country to country. They use data from the Encyclopedia Britannica and other sources to construct three separate measures of fractionalization. The first combines racial and linguistic characteristics (ETHNIC), the second is based on data for shares of languages spoken as "mother tongues"

(LANG), and the third describes differences in religion (RELIG). I employ all four measures of fractionalization in the tree regressions in ways made clear below.

Table 1 provides a detailed description as well as summary statistics for proxies for the three classes of fundamental determinants discussed above.

5. Results

Since GUIDE selects threshold variables with negligible bias, I include in my benchmark regression trees for both development clubs and growth regimes all nine fundamental determinant variables. To affirm the robustness of the benchmark tree structures, I further employ eight different combinations of fundamental determinant variables for each series of regression exercises; i.e., for development clubs and for growth regimes. These eight additional models (Models 1-8) and the benchmark (Model 0) are described in Table 2. All eight models include the proxy for geographic isolation (LCR100KM). Differences across models are due to different configurations of climate, institutions, and fractionalization variables. Specifically, I interchange the two climate variables as well as the two quality of institutions variables, and use ELF60 and Alesina et al.'s set of ethnic, linguistic, and religious fractionalization variables as substitutes for each other across these models.

We hope to observe two forms of robustness by using different combinations of threshold variables. First, we want to affirm that threshold variable selection by GUIDE is in fact unbiased. We expect that the uncovered structure should not vary dramatically across models employing different combinations of proxies for fundamental determinants. Finally, we want to see if the results are robust to small changes in the data set generated by the inclusion or exclusion of countries due to variations in missing values across variables. The effective number of observations for the development clubs and growth regimes exercises after eliminating rows with missing values range from 84 to 98 for the former and from 150 to 160 for the latter. As mentioned above, we also report 95% confidence intervals for threshold value estimates using Hansen's methodology. While the tree structures generated by GUIDE for development clubs and growth regimes offer us an interpretable relationship between fundamental determinants and economic performance, the confidence bounds provide us with a measure of the uncertainty over the classification of particular countries into each club or regime.

An important caveat to the findings is the difficulty in dealing with issues of endogeneity in the sample splitting context. Caner and Hansen (2004) consider the case of endogenous (slope) variables in the regression equation, but their method does not extend to the case of endogenous threshold variables for which there is currently no solution in the (econometrics) literature. This is potentially problematic since institutions, in particular, have been argued to be potentially endogenous. Nevertheless, we should note that attempts to deal with the problem of the endogeneity of institutions in the linear context are far from satisfactory. The most credible instruments (proposed by Acemoglu et al., 2001), severely restricts the number of cross-country observations (to just the set of ex-European colonies). More seriously, Brock and Durlauf have argued against the credibility of *any* exclusion restrictions given the problem of model uncertainty (Brock and Durlauf, 2001; page 12). I do not wish to overstate any claims in this paper and suggest that the findings be viewed as uncovering patterns in the data rather than asserting statements of causality. However, given that problems of endogeneity plague virtually all studies in the empirical growth literature, I also suggest that these findings may nevertheless be qualitatively no better or worse than those which characterize the broader literature.

5.1 *Results for Development Clubs*

The regression tree results provide robust support for the existence of three development clubs. Figure 2(a) provides a schematic characterization of these development clubs. There are two main findings. The first is that the quality of institutions, and, in particular, their interaction with ethnic fractionalization, is the robust determinants that drive long-run differences in economic development. That is, the data rejects linearity in favor of multiple (two) development thresholds. For the group of countries with quality of institutions above a threshold value (EXPROP8497>8.40), ethnic fractionalization has no impact on development whereas for countries that have quality of institutions below the threshold, having levels of ethnic fractionalization above a threshold value (ELF60>0.605) is associated with substantially lower levels of development.

The predicted level of real per capita GDP for the high quality institutions group of countries is almost 14 times that of the low quality institutions-high ethnic fractionalization group. Further, the low quality institutions-low ethnic fractionalization group enjoys three times the predicted income of the group with high ethnic fractionalization. Interestingly, religious pluralism at least as measured by Alesina et al.'s religious fractionalization variable (RELIG) does not appear to have an important role in explaining divergence.

The second finding is that geography plays no role in explaining divergence in long-run levels of development once institutions and fractionalization are controlled for. Neither factors related to climate nor those related to geographic isolation are identified by the regression trees as being important threshold variables. The results lend partial support therefore to work such as Rodrik et al. (2004) that suggest that once institutions are controlled for, factors such as climate and openness (which we relate to geographic accessibility in this study) are no longer significant determinants of development outcomes.

These results are robust to variation in the set of candidate threshold variables. Table 3 describes, for each of the eight combination of fundamental determinant variables (Models 1-8) discussed above, the partitioning of the joint support of these variables by the regression tree to obtain development clubs. It also lists the predicted per capita income levels for countries in each development club. The results are fully consistent with those we obtained for our benchmark model.

I now organize the discussion of my results in the context of regions of the world. Table 5 gives the breakdown of countries by development clubs and regions for each of the tree regression results (for Models 0 to 8) shown in Table 3. The number in each cell in Table 5 identifies the corresponding country as belonging to the development club of the same number in Table 3. Figure 2(b) shows Hansen's 95% confidence intervals for the first and second threshold splits for the benchmark model (see Figure 2(a)).

Unsurprisingly, western industrialized economies (Western European countries and their offshoots) are consistently classified as members of the development club with the highest level of predicted per capita income[11]. We are also confident that these countries are correctly classified into this group since they register above the upper bound of the 95% confidence interval for the quality of institutions (EXPROP8497) threshold. It should be noted, however, that some countries in this group also have relatively high levels of ethnic fractionalization. For instance, both Easterly and Levine's ethno-linguistic index (ELF60) and Alesina et al.'s ETHNIC variable place Canada's degree of ethnic fractionalization at levels higher than those of some Sub-Saharan African countries (e.g., Ethiopia, Malawi, and Mali). Other countries like Belgium, Switzerland, and the US, have ethnic

[11] The exception is Greece which occasionally gets classified along with members of the second highest predicted income club.

fractionalization levels that are close to the threshold value for the low quality institutions-high ethnic fractionalization development club. Western industrialized countries, however, generally have the highest quality institutions compared to the rest of the world. The result therefore lends support to the view that better quality institutions potentially mitigate any negative effects from higher levels of diversity.

The picture for Asia is more complex. A handful of Asian countries, Hong Kong, Japan, the Republic of Korea[12], and Singapore (as well as Taiwan) routinely belong to the development club with the highest predicted income. They also come in above the 95% confidence interval for the quality of institutions threshold. However, the majority of these are essentially small island states. Other Asian countries like India, Indonesia, Pakistan, and the Philippines tend to fall into the development club with the lowest predicted income (the low quality institutions-high ethnic fractionalization club). While there may be some uncertainty over whether the two largest of these countries (India and Indonesia) were correctly classified, the 95% confidence bounds for EXPROP8497 suggest that Pakistan and the Philippines were correctly classified as belonging to that group.

Malaysia, however, presents itself as a particularly interesting case. Malaysia frequently finds itself in the development club with either the highest or the next highest level of predicted income despite its high level of ethnic fractionalization. With values for ELF60 and ETHNIC at 0.79 and 0.59 respectively, Malaysia's level of ethnic fractionalization either dominates or compares with those of poor Asian countries[13]. In fact, for models where ELF60 was used, if the quality of

[12] The Republic of Korea is marginally displaced into the next richest development club for model 1. However, it should be noted that the Republic of Korea tends to enter lower in quality of institutions ratings not through its own failings, but because of the risks to property rights posed by possible actions from its northern neighbor.

[13] Values for ELF60 and ETHNIC for India, Indonesia, Pakistan, and the Philippines are, respectively, 0.89 and 0.82, 0.76 and 0.74, 0.64 and 0.71, and, 0.74 and 0.24.

Malaysia's institutions had been low enough, its high level of ethnic fractionalization would have placed it in the development club with the lowest predicted income (as is the case with our benchmark). This holds out hope for developing countries that good institutions are potentially able to overcome disadvantages posed by a more diverse population.

Turning now to the Middle East and North Africa, we see that these countries are not plagued with high levels of diversity. However, their low institutional quality places them in the middle-income (low quality institutions-low ethnic fractionalization) development club.

A similar situation applies to countries in Latin America and the Caribbean. Here again, the story is clearly one of institutions. With only a few exceptions like Guatemala and Paraguay, ethnic fractionalization is generally low (in the data) for this group of countries. What occasionally separates countries like Brazil, Chile, and Costa Rica from the other Latin American and Caribbean countries are institutions of moderately higher quality. In fact, the data suggests a few other countries in this region have institutions that are close to the threshold value (in particular, Venezuela and Mexico). This holds out hope that policies targeted at improving institutions would result in sizeable pay-offs for these countries. Table 3 indicates that the predicted per capita income for the high quality institutions development club is, on average, roughly 4 times that that of the low quality institutions-low ethnic fractionalization development club.

Finally, let us turn to Sub-Saharan Africa. This region suffers from a confluence of negative factors from having extreme low levels of institutional quality, some of the highest levels of ethnic fractionalization, and a landlocked interior. These problems are to a large extent part of the sub-continent's colonial legacy. Present day African political boundaries did not arise naturally but instead were arbitrarily drawn up during de-colonization. In many cases these artificial borders encompassed groups of people with little in common culturally or sociologically[14].

The results highlight the influence of Africa's past on its future. With few exceptions, notably Botswana, Madagascar, South Africa, and Zimbabwe, countries in Sub-Saharan Africa are consistently classified into the group of countries with the lowest predicted income levels (the low quality institutions-high ethnic fractionalization development club). And, of the exceptions, the 95% confidence bounds for EXPROP8497 indicate that only Botswana may have in fact been a high predicted income country (i.e., a group 3 country) that was misclassified. In fact, the confidence bounds for EXPROP8497 and ELF60 together provide strong evidence that low-quality institutions coupled with high ethnic fractionalization explains the dismal performance of Sub-Saharan Africa as a region.

What is particularly surprising is that there is no evidence to suggest that climate-related factors are at the root of Africa's under-development. This is in contrast to a large body of work in the literature suggesting such a link. The findings here do not support the conclusion that "the tropics are damned not just, or even mainly, by bad policies, but by difficult inherent conditions [Gallup, Sachs, and Mellinger, 1999]". In fact, the results in this paper point strongly to the possibility that Africa's ills are fundamentally socio-political in nature. In particular, the results agree with those of Easterly and Levine (1997) citing the importance of high levels of ethnic fractionalization in explaining Africa's poor performance.

5.2 *Results for Growth Regimes*

---

[14] In Chad, Kenya, Nigeria, and Sudan, for instance, different ethic groups were simply forced to co-exist. Conversely, in southwest Ivory Coast, the Cavalla River separates the Kru and Grebo tribes from their counterparts in Liberia.

Table 4(a) details growth regimes for the benchmark (Model 0) and also for tree regressions utilizing each of the eight combinations of fundamental variables (Models 1-8). I find evidence for the existence of two growth regimes. The results, in this case, suggest that institutions are of first-order importance in determining economic outcomes. Out of the nine trees, seven, including the benchmark, routinely separate a set of high quality institution countries from the rest of the world. Table 7 provides the regression estimates for the nine models as well as ordinary least squares estimates for the standard (unconstrained) Mankiw, Romer, and Weil (MRW) growth regression with homogenous coefficients. The country breakdowns for each model are shown in Table 6.

From the tree regression results in Table 7, we see that there is substantial parameter heterogeneity across the two growth regimes. Population growth appears to have a much larger negative impact on low quality institutions countries than high quality institutions ones. The coefficient to log population growth (MNGD) is around twice as negative for countries in the former regime compared to the latter. The coefficient for log investment share (MINV) is also much larger for the high quality institutions regime compared to the low quality institutions regime (also around twice as large). There is more ambiguity over the coefficient for log schooling (MSCH25) with some models suggesting that schooling is more important to low quality institutions countries (Models 1 and 3) and others, including the benchmark, that it is more important to countries with high quality institutions (Models 0, 4, 6, 7 and 8).

The 95% confidence bounds for the threshold split (see Figure 3(b)) suggest that the classification of countries into the two regimes exhibited some uncertainty but was nevertheless reasonably accurate; a quarter of all countries fell within the bounds. Three features are evident when comparing the country breakdowns for growth regimes (Table 6) with those for development clubs (Table 5). First, we see that countries that were classified as members of the development club with the highest predicted income (high quality institutions group) all conform to the good quality institutions growth regime. Second, countries that were classified as members of the development club with the lowest predicted income (low quality institutions-high ethnic fractionalization group) remained predominantly within the low quality institutions growth regime. The exceptions comprised a handful of Asian countries (India, Indonesia, Malaysia, and Thailand) that went from the lowest predicted income development club to the high quality institutions growth regime. With few exceptions, Sub-Saharan African countries were classified into the low quality institutions regime. Finally, members of the middle income development club (low quality institutions-low ethnic fractionalization group) were distributed primarily into the low quality institutions regime with the exception of Brazil, Chile, Columbia, Mexico, Papua New Guinea, Turkey, and Botswana.

There are at least two explanations for the disparity between the number of development clubs and the number of growth regimes. The explanations hinge to a large extent on the structural interpretation of the various regressions trees. If we view the development clubs regression as primarily an exploratory exercise to uncover patterns in the data, then the reason for the disparity in groupings may be attributed to the possibility that countries in the low quality institutions regime are simply not close enough to the long-run steady state distribution for the development club groupings to converge into agreement with the growth regimes.

There is some evidence for this view. The estimates for the coefficient to initial income (MGDP0) for the low quality institutions growth regime are significant[15], negative, and about a third in absolute value of those for the high quality institutions regime. This suggests that the rate of convergence is substantially slower for low quality institutions countries compared to those with high quality institutions. The large number of development clubs may, therefore, simply reflect

[15] At the 10% level for our benchmark model and for Model 6, and at the 5% level for the other seven remaining models.

transient heterogeneity that will be resolved in the long-run. That is, our results suggest a polarization of the middle-class in the long-run to conform to the findings for two growth regimes. This story is consistent with findings of "twin peak"-ness in studies on the evolution of the distribution of cross-country income (see, for example, Quah, 1997). As we noted above, we find like those studies evidence for persistence; since all rich countries remain in the high quality institutions growth regime while most poor countries remain in the low quality growth regimes, for growth "miracles"; i.e., those countries going from the low and middle income development club to the high quality growth regime, and for growth "disasters"; i.e., those countries going from the middle income development club to the low quality growth regime.

On the other hand, if we maintain that the development clubs uncovered by our tree regressions do not simply reveal patterns in the data, but are, in fact, delivering long-run predictions for levels of development, then the difference between the number of growth regimes and development clubs may suggest that the MRW specification does better at describing the growth processes of industrialized economies than less developed ones. In particular, the dissipation of countries in the low quality institutions regime into multiple development clubs suggests in this case that there exists (residual) heterogeneity in the growth process of less developed countries that may not be well-captured by the MRW regression specification. If this is true, then more attention needs to be paid to characterizing growth processes for developing countries.

6. Conclusion

This paper attempts to provide a deeper understanding of how fundamental determinants interact to hinder or facilitate development outcomes for different groups of countries. Such a view is not possible with mono-causal approaches that ignore the possibility of heterogeneity in economic processes. I find that institutions play a central role in accounting for the divergence in cross-country growth experiences. The results for development clubs point further to a key role for ethnic fractionalization in hindering long-run development. The findings suggest that more effort needs to be made to understand the connection between ethnic fractionalization and institutional quality, and their impact on development. The preliminary analysis suggests that sufficiently high quality institutions may be necessary if the negative impact on economic outcomes from higher levels of ethnic fractionalization is to be mitigated. Finally, I find no role for geography in hindering the development of Sub-Saharan African countries. While one has to be extremely cautious in extracting policy implications from growth empirics, the results in this paper would tend to lend support to the view that Africa would be best served by improved institutions and policies to promote nation-building across ethnic lines rather than policies aimed at mitigating the effects of the physical environment. There are simply no "magic bullet" solutions to Africa's problems.

References

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

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

Alesina A, Baqir R, Easterly W. 1999. Public Goods and Ethnic Divisions. *Quarterly Journal of Economics* 114(4):1243-1284.

Alesina, A, Devleeschauwer A, Easterly W, Kurlat S, Wacziarg R. 2003. Fractionalization. *Journal of Economic Growth* 8:155-194.

Alesina A, La Ferrara E. 2002. Who Trusts Others? *Journal of Public Economics* 85:207-234.

Barro R J. 1996. Democracy and Growth. *Journal of Economic Growth* 1(1): 1-27.

Barro, R J, Lee J.-W. 2000. International Data on Educational Attainment: Updates and Implications. Mimeo. Center for International Development, Harvard University.

Bloom D E, Sachs J. 1998. Geography, Demography, and Economic Growth in Africa. *Brookings Papers on Economic Activity* 2:207-295.

Bloom D E, Canning D, Sevilla J. 2003. Geography and Poverty Traps. *Journal of Economic Growth* 8:355-378.

Breiman L, Friedman J H, Olsen R A, Stone C J. 1984. Classification and Regression Trees, Wadsworth, Belmont.

Brock W A. 2001. Complexity-based Models in Cycles and Growth: Any Potential Value? in Cycles Growth and Structural Change: Theories and Empirical Evidence, L. F. Punzo, editor, Routledge, London, 301-338.

Brock W A, Durlauf S N. 2001. Growth Empirics and Reality. *World Bank Economic Review* 15(2):229-272.

Caner M, Hansen B E. 2004. Instrumental Variable Estimation of a Threshold Model. *Econometric Theory* 20:813-843.

Canova F. 2004. Testing for Convergence Clubs in Income Per Capita: A Predictive Density Approach. *International Economic Review* 45(1):49-77.

Chamon M, Kremer M R. 2006. Asian Growth and African Development. *American Economic Review* 96(2):400-404.

Desdoigts A. 1999. Patterns of Economic Development and the Formation of Clubs. *Journal of Economic Growth* 4 (3):305-30.

Diamond J. 1997. Guns, Germs and Steel: The Fates of Human Societies, W. W. Norton, NY.

Doyle P. 1973. The Use of Automatic Interaction Detector and Similar Search Procedures. *Operational Research Quarterly* 24:465-467.

Durlauf S N, Johnson P A. 1995. Multiple Regimes and Cross Country Behavior. *Journal of Applied Econometrics* 10 (4):365-384.

Durlauf S N, Kourtellos A, Minkin A. 2001. The Local Solow Growth Model. *European Economic Review* 45:4-6.

Durlauf S N, Kourtellos A, Tan C M. 2007. Are Any Growth Theories Robust. *Economic Journal* forthcoming.

Easterly W, Levine R. 2003. Tropics, Germs, and Crops: How Endowments Influence Economic Development. *Journal of Monetary Economics* 50(1):3-39.

Easterly W, Levine R. 1997. Africa's Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics* 112(4):1203-50.

Frankel J, Romer D. 1999. Does Trade Cause Growth? *American Economic Review* 89 (3):379-399.

Gallup J, Sachs J, Mellinger A. 1999. Geography and Economic Development. *International Science Review* 22(2):179-232.

Galor O. 2005. From Stagnation to Growth: Unified Growth Theory in Handbook of Economic Growth, P. Aghion and S. Durlauf, eds., Amsterdam: North Holland, forthcoming.

Galor O, Moav O. 2002. Natural Selection and the Origin of Economic Growth. *Quarterly Journal of Economics* 117:1133-1191.

Glaeser E L, Laibson D, Sheinkman J A, Soutter C L. 2000. Measuring Trust. *Quarterly Journal of Economics* 65:811-846.

Gonzalo J, Pitarakis J-Y. 2002. Estimation and Model Selection Based Inference in Single and Multiple Threshold Models. *Journal of Econometrics* 110:319-352.

Hall R, Jones C. 1999. Why Do Some Countries Produce So Much More Output per Worker than Others? *Quarterly Journal of Economics* 114(1):83-116.

Hansen B E. 2000. Sample Splitting and Threshold Estimation. *Econometrica* 68:575-603.

Hansen B E. 1999. Threshold Effects in Non-Dynamic Panels: Estimation, Testing and Inference. *Journal of Econometrics* 93:345-368.

Hastie T, Tibshirani R, Friedman J H. 2001. The Elements of Statistical Learning Springer-Verlag.

Heston A, Summers R, Aten B. 2002. Penn World Table Version 6.1 Center for International Comparisons at the University of Pennsylvania (CICUP).

Knack S, Keefer P. 1995. Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures. *Economics and Politics* 7(3):207-227.

La Porta R, Lopez-de-Silanes F, Shleifer A, Vishny R. 1999. The Quality of Government. *Journal of Law, Economics and Organization* 15(1):222-79.

Loh W-Y. 2002. Regression Trees with Unbiased Variable Selection and Interaction Detection. *Statistica Sinica* 12:361-386.

Mankiw N G, Romer D, Weil D N. 1992. A Contribution to the Empirics of Economic Growth. *Quarterly Journal of Economics* 107(2):407-437.

Masanjala W, Papageorgiou C. 2007. Rough and Lonely Road to Prosperity: A Reexamination of the Sources of Growth in Africa using Bayesian Model Averaging. *Journal of Applied Econometrics* forthcoming.

Masters W, McMillan M. 2001. Climate and Scale in Economic Growth. *Journal of Economic Growth* 6(3):167-86.

Masters W, McMillan M. 2003. An African Growth Trap: Production Technology and the Time-Consistency of Agricultural Taxation, R&D and Investment. *Review of Development Economics* 7(3):179-91.

McArthur J, Sachs J. 2001. Institutions and Geography: Comment on Acemoglu, Johnson and Robinson (2000). NBER Working Paper No. 8114.

McMillan M. 2001. Why Kill the Golden Goose? A Political Economy Model of Export Taxation. *Review of Economics and Statistics* 83(1):170-184.

Persson T, Tabellini G. 1994. Is Inequality Harmful for Growth? *American Economic Review* 84(3):600-621.

Quah D. 1997. Empirics for Growth and Distribution: Polarization, Stratification, and Convergence Clubs. *Journal of Economic Growth* 2: 27-59.

Radelet S, Sachs J. 1998. Shipping Costs, Manufactured Exports, and Economic Growth. HIID, mimeo.

Rodrik D, Subramanian A, Trebbi F. 2004. Institutions Rule: The Primacy of Institutions Over Geography and Integration in Economic Development. *Journal of Economic Growth* 9(2):131-165.

Sachs J. 2003. Institutions Don't Rule: Direct Effects of Geography on Per Capita Income. NBER Working Paper No. 9490.

Solow R. 1956. A Contribution to the Theory of Economic Growth. *Quarterly Journal of Economics* 70(1):65-94.

Technical Appendix

This section reproduces the key algorithms used in the GUIDE tree regression software. Loh (2002) provides more details about GUIDE as well as a description of other tree generating options within GUIDE.

First, define the following classes of covariate variables:

- *n*-variable: a numerical-valued predictor used to fit the terminal node regression model and to split the nodes in the tree;
- *f*-variable: a numerical-valued predictor used to fit the terminal node regression model but not to split the nodes in the tree;
- *s*-variable: a numerical-valued predictor used to split the nodes in the tree but not to fit the terminal node regression model;
- *c*-variable: a categorical predictor used to split the nodes in the tree but not to fit the terminal node regression model.

The first two algorithms determine the choice of the splitting variable at each node of the tree.

Algorithm 1: Chi-square tests for linear fit.

1. Obtain the residuals from a linear model fitted to the *n*- and *f*-variables, leaving out the *s*- and *c*-variables.
2. For each *n*-variable, divide the data into four groups at the sample quartile; construct a 2 x 4 contingency table with the signs of the residuals (positive versus non-positive) as rows and the groups as columns; count the number of observations in each cell and compute the χ^2-statistic and its theoretical p-value from a χ_3^2 distribution.
3. Do the same for each *s*- and *c*-variable. For the latter, the categories of the variable form the columns of the table. Columns with zero column totals are omitted.
4. To detect interactions between each pair of *n*-variables (X_i, X_j), divide the (X_i, X_j)-space into four quadrants by splitting the range of each variable into two halves at the sample median; construct a 2 x 4 contingency table using the residual signs as rows and the quadrants as columns; compute the χ^2-statistic and p-value. Again, columns with zero column totals are omitted.
5. Do the same for each pair of *s*-variables.
6. Also do the same for each pair of *c*-variables using their value pairs to divide the sample space. For example, if X_i and X_j take c_i and c_j unique values, respectively, the χ^2-statistic and p-value are computed from a table with 2 rows and number of columns equal to $c_i c_j$ less the number of zero columns.

7. Compute a χ^2-statistic and p-value for each pair $\left(X_i, X_j\right)$ where X_i is an *n*-variable and X_j is a *c*-variable. If X_j has *c* categories, the table has 2 rows and number of columns equal to 2*c* less the number of zero columns.
8. Similarly, compute a χ^2-statistic and p-value for each pair $\left(X_i, X_j\right)$ where X_i is an *s*-variable and X_j is a *c*-variable.
9. Finally, do the same for each pair where X is an *s*-variable and X_j is an *n*-variable as in step 4.

Algorithm 2: *Choosing the splitting variable.*

1. Note that 9 sets of Chi-square tests are computed: 3 sets to detect curvature in the *n*-, *s*-, and *c*-variables, 3 sets to detect interactions between pairs of variables of the same type, and 3 sets to detect interactions between pairs of predictors of different types.
2. If the smallest p-value comes from a curvature test, the associated variable is selected to split the node.
3. Suppose instead that a pair of variables is selected because their interaction test is the most significant among the curvature and interaction tests.
4. If neither is a *n*-variable, choose the one with the smaller curvature p-value.
5. If both are *n*-variables, temporarily split the node along the sample mean of each variable; choose the variable whose split yields the smaller total SSE.
6. If exactly one is an *n*-variable, choose the other variable.

Once a splitting variable (call it X_j) has been chosen, we need to determine the split value for that variable. This is done in the next algorithm.

Algorithm 3: *Choosing the split value.*

1. Consider the two partitions of the sample space Y x X defined as follows,

$$A_1^j(s) = \left\{\left(y_i, X_i\right) \in Y \times X \mid x_j^i \leq s\right\}$$

$$A_2^j(s) = \left\{\left(y_i, X_i\right) \in Y \times X \mid x_j^i > s\right\}$$

where $i = 1, \ldots, n$ indexes observations and $x_j^i \in X_j$ for all *i*.

2. The task is to determine the value s.

3. Let $\hat{\beta}^{1}_{(j,s)}$ be the OLS estimator of the regression of Y on X for the subset of observations that conforms to the partition $A_1^j(s)$, and $\hat{\beta}^{2}_{(j,s)}$ be the OLS estimator of the regression of Y on X for observations conforming to partition $A_2^j(s)$.

4. Find the split value s for splitting variable X_j that minimizes the sum of squared residuals (SSR):

$$\frac{1}{n_1}\sum_{(y_i,X_i)\in A_1^j(s)}\left(y_i - X_i\hat{\beta}^{1}_{(j,s)}\right)^2 + \frac{1}{n_2}\sum_{(y_i,X_i)\in A_2^j(s)}\left(y_i - X_i\hat{\beta}^{2}_{(j,s)}\right)^2$$

To grow a tree, therefore, GUIDE starts with the set of all observations and applies the three algorithms above to find a splitting variable and a split value leaving two mutually exclusive subsets of observations the union of which forms the set of all observations. It then continues to apply this same procedure to each of the resultant subsets, and then to the subsets of observations resulting from those, and so forth iteratively until the number of observations in the subset falls below a certain predetermined value. In our exercises, we take this minimum number of observations to be the default value set by GUIDE. After the tree is grown, it has to be "pruned" in order to avoid over-fitting the data. This is done using Cost Complexity pruning.

Algorithm 4: *Cost Complexity Pruning*

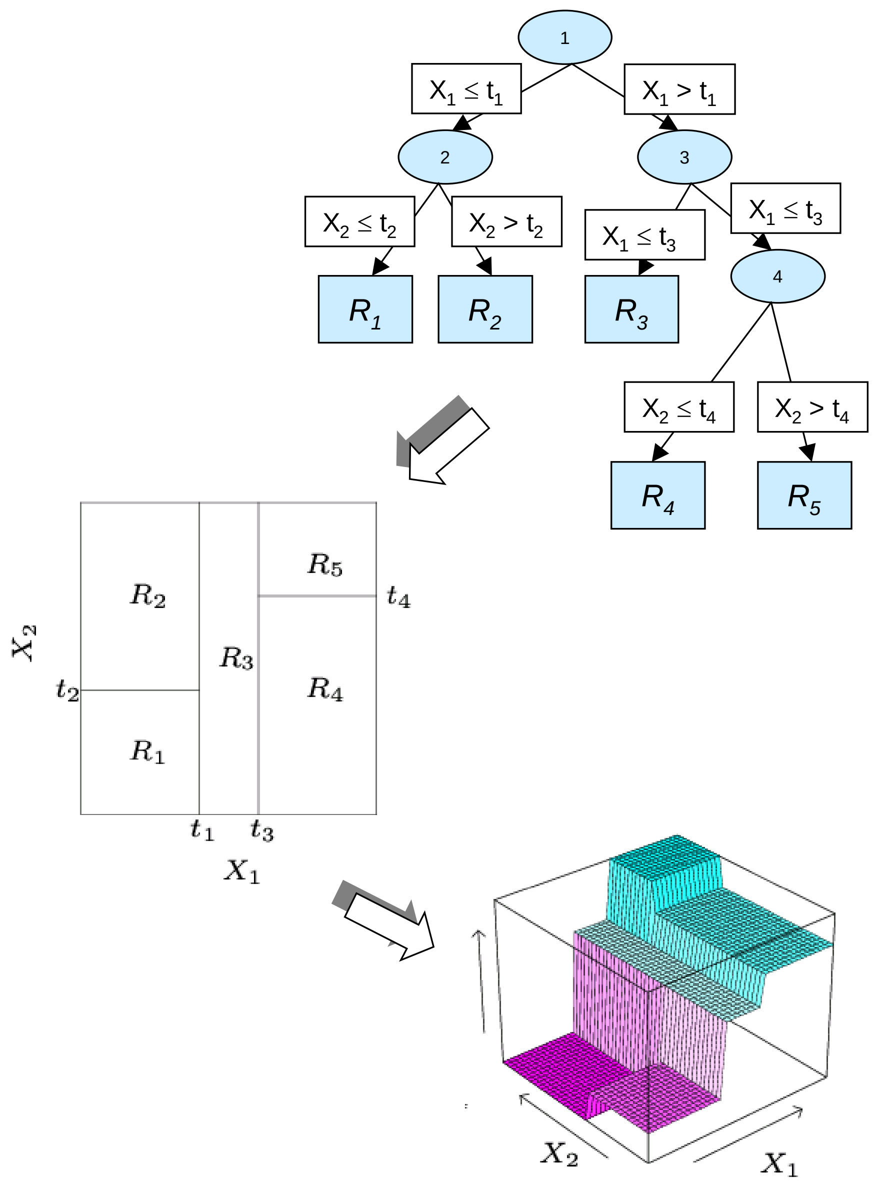
1. First denote the fully-grown tree to be pruned as T_0. Now, consider a sub-tree $T \prec T_0$ with *b* terminal nodes.
2. Define the Cost Complexity criterion for given penalty parameter α (as applied to sub-tree T) as,

$$C_\alpha(T) = \sum_{m=1}^{b}\sum_{(y_i,X_i)\in TN(m)}\left(y_i - X_i\hat{\beta}_m\right)^2 + \alpha\cdot b$$

3. The idea behind the use of the Cost Complexity criterion is to attempt to minimize SSR but to penalize over-fitting through the use of an overly complex tree.
4. Denote the tree for which $C_\alpha(T)$ is minimized (for given α) as $T_\alpha \prec T_0$.
5. The SSR for T_α is obtained by *V*-fold cross-validation.
6. Locate the tree for which α minimizes the cross-validated SSR.
7. Choose the smallest tree within one standard error of the smallest cross-validated SSR (this could be the tree that minimizes the cross-validated SSR).

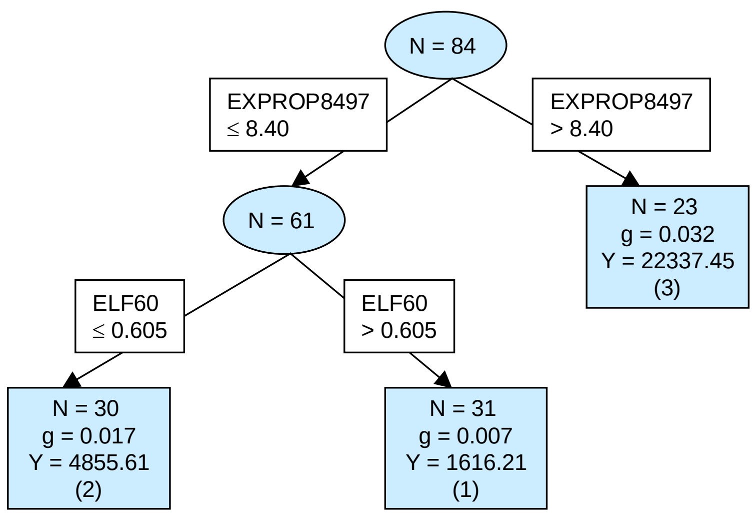
For the GUIDE generated trees reported in this paper, *V* was set to the maximum value available; i.e., the number of observations for each sample.

Figure 1*: Simple Tree Schematic



*Schematic due to Hastie, Tibshirani, and Friedman (2001).

Figure 2(a)[*]: Tree Diagram Showing Development Clubs



[*]Note that N gives the number of observations in each node, g stands for the average growth rate in real per capita GDP from 1960-99 for countries in the terminal node, and Y is the level of real per capita GDP predicted by the tree regression.

Figure 2(b): Hansen (2000) Confidence Intervals for Development Clubs

First Split:

Threshold Variable	EXPROP8497
Threshold Estimate	8.3846154
95% Confidence Interval:	[7.000000, 8.384615]

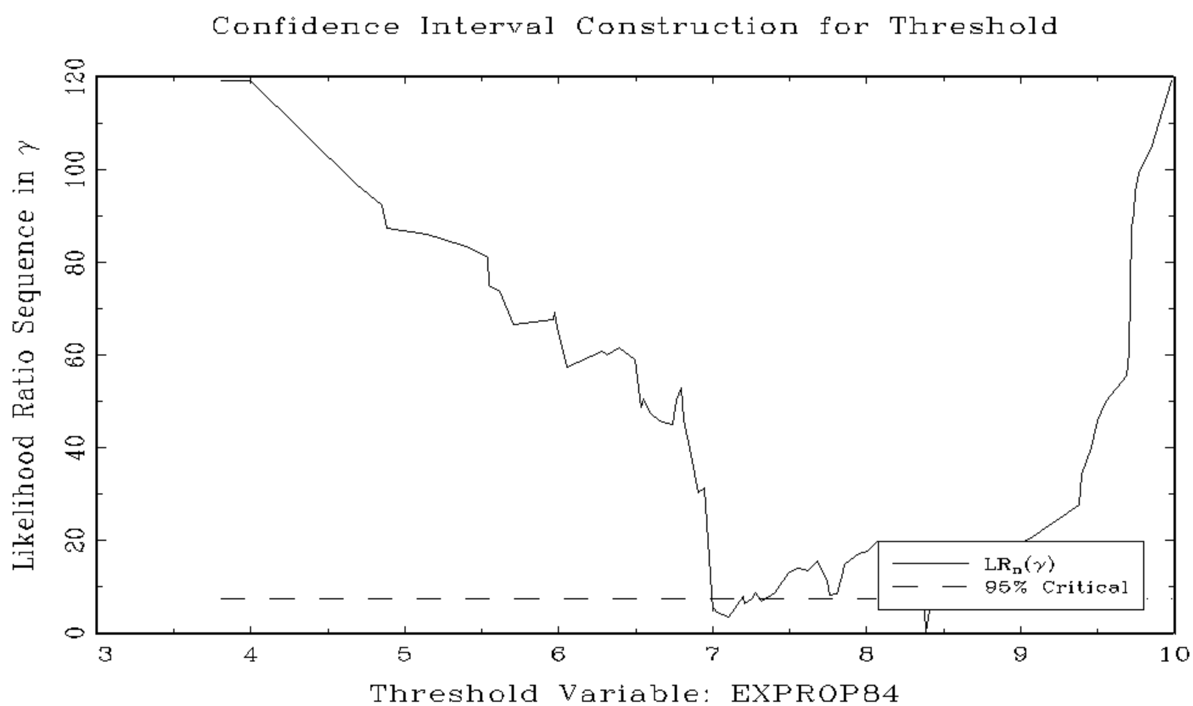


Figure 2(b): Hansen (2000) Confidence Intervals for Development Clubs (cont.)

<u>Second Split:</u>

Threshold Variable	ELF60
Threshold Estimate	0.58999997
95% Confidence Interval:	[0.58999997, 0.58999997]

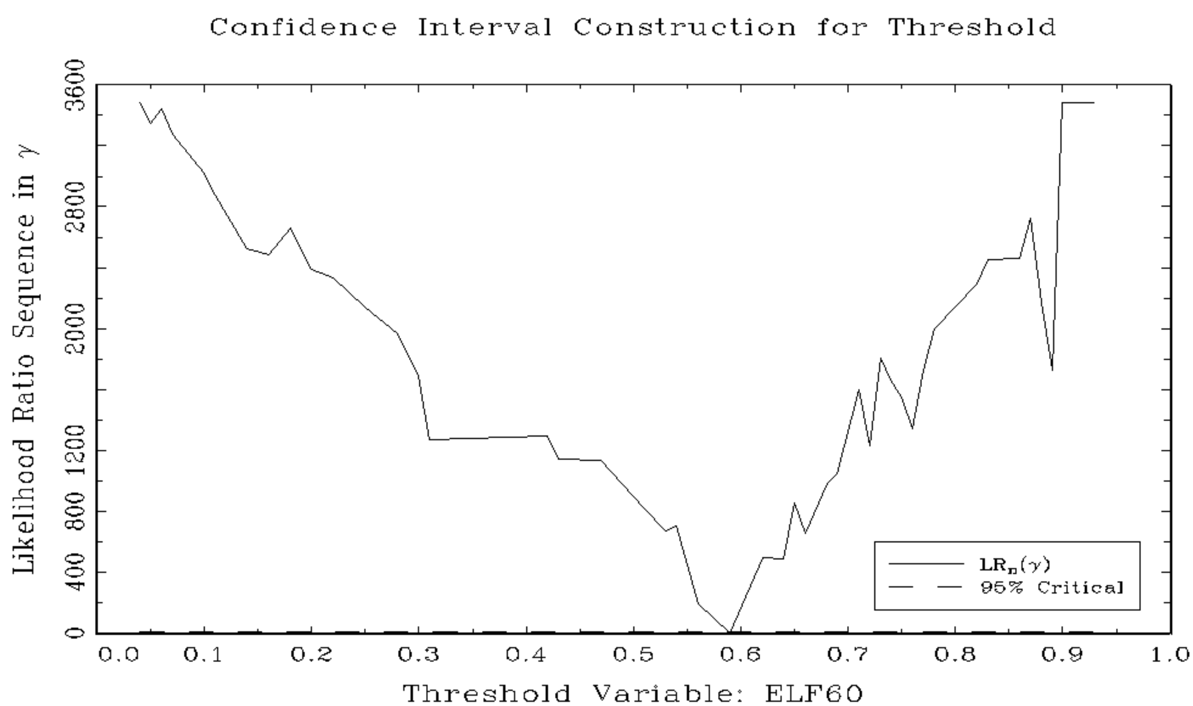
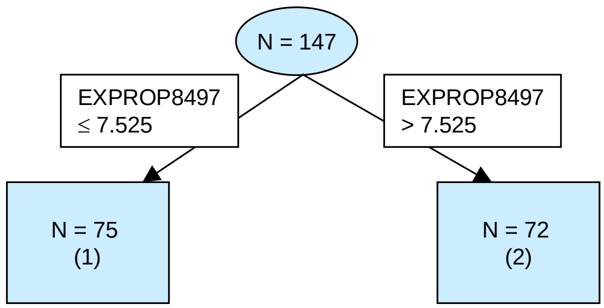


Figure 3(a)*: Tree Diagram Showing Multiple Growth Regimes



*Note that N gives the number of observations in each node. Please refer to Table 4 for the list of countries for each growth regime, and to Table 7 for the tree regression estimates of coefficients to neoclassical growth proximates.

Figure 3(b): Hansen (2000) Confidence Intervals for Multiple Regimes

Threshold Variable	EXPROP8497
Threshold Estimate	7.4928571
95% Confidence Interval:	[7.000000, 7.950000]

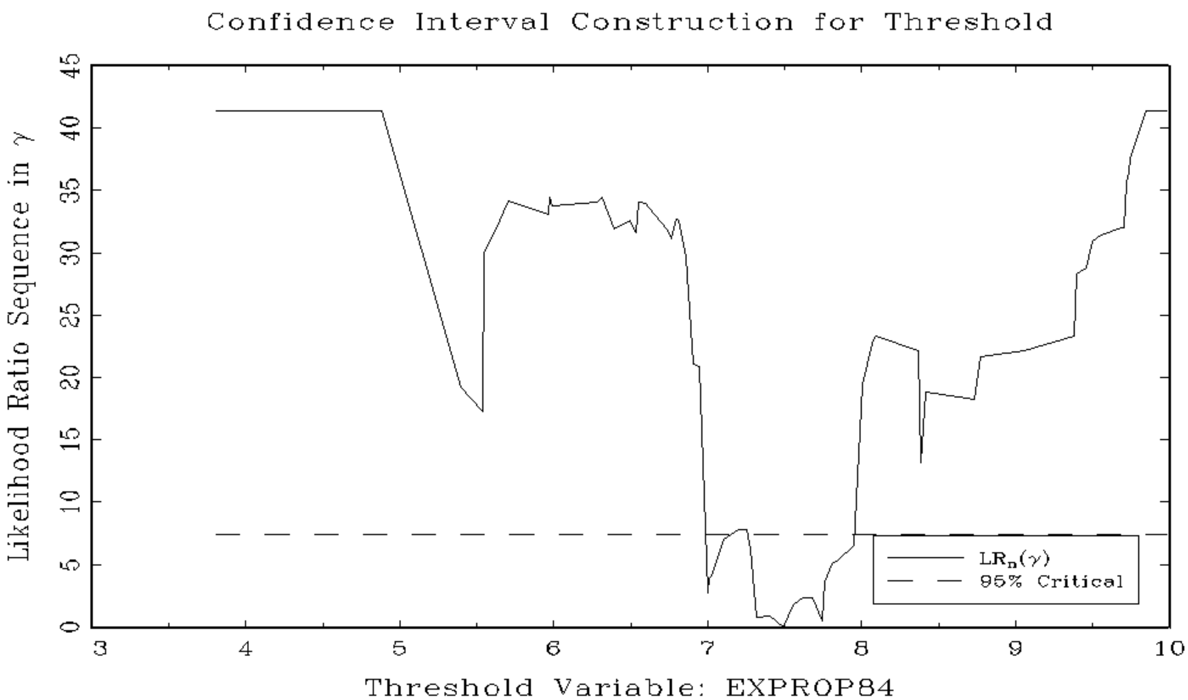


Table 1[#]: Summary Statistics

Variables	EXPROP8497	ICRG8497	FROST5	ZTROPICS	LCR100KM	ELF60	ETHNIC	LANG	RELIG
Min.	3.807	2.342	0.000	0.000	0.000	0.000	0.000	0.002	0.002
Max.	10.000	9.957	1.000	1.000	94.000	0.930	0.930	0.923	0.860
Mean	7.417	6.286	0.439	0.167	1.101	0.424	0.434	0.391	0.437
Median	7.300	5.855	0.141	0.000	0.392	0.430	0.418	0.363	0.460
Std. Dev.	1.604	1.942	0.462	0.259	7.646	0.298	0.254	0.286	0.234

[#] Description of Fundamental Determinant Variables:

1. **FROST5**: Proportion of a country's land area that experiences more than 5 frost-days per month in winter.
2. **ZTROPICS**: Percentage of a country's land area that is classified as a tropical eco-zone.
3. **LCR100KM**: Percentage of a country's land area within 100 km of ice-free coast. Calculated using 100 km from ice-free coast or navigable river buffer (created in ArcView using Plate Caree equidistant projection).
4. **EXPROP8497**: Risk of Expropriation of Private Investment; annual average calculated over the years ranging 1984-97. This variable evaluates the risk "outright confiscation and forced nationalization" of property. Lower ratings "are given to countries where expropriation of private foreign investment is a likely event."
5. **ICRG8497**: International Country Risk Guide Index; annual average calculated over the years ranging 1984-97. This is a broad measure of institutional quality which aggregates across five variables measuring (1) the quality of the bureaucracy, (2) corruption in government, (3) rule of law, (4) expropriation risk, and (5) repudiation of contracts by government. Higher values of ICRG8497 correspond to better quality of institutions.
6. **ELF60**: Index of ethno-linguistic fractionalization, 1960. Measures the probability that two randomly selected people from a given country will not belong to the same ethno-linguistic group. Atlas Narodov Mira (1964).
7. **ETHNIC**: Measure of ethnic fractionalization from data which combines racial and linguistic characteristics (based on data from the Encyclopedia Britannica (2001) and other sources).
8. **LANG**: Measure of linguistic fractionalization based on data describing shares of languages spoken as "mother tongues" (based on data from the Encyclopedia Britannica (2001)).
9. **RELIG**: Measure of religious fractionalization (based on data from the Encyclopedia Britannica (2001)).

Table 2[◊]**: Model Specifications**

Model	Institutions		Geography			Fractionalization			
	EXPROP8497	*ICRG8497*	*FROST5*	*ZTROPICS*	*LCR100KM*	*ELF60*	*ETHNIC*	*LANG*	*RELIG*
0	X	X	X	X	X	X	X	X	X
1		X		X	X	X			
2		X	X		X	X			
3		X		X	X		X	X	X
4		X	X		X		X	X	X
5	X			X	X	X			
6	X		X		X	X			
7	X			X	X		X	X	X
8	X		X		X		X	X	X

[◊] All fundamental determinant variables were included in the set of possible threshold variables in the benchmark tree regression for development clubs and growth regimes (labeled as Model 0). To test for robustness, eight additional specifications (Models 1-8) were employed. In particular, all eight models include the proxy for geographic isolation (LCR100KM). Differences across models are due to different configurations of institutions and fractionalization variables.

Table 3: Regression Trees for Development Clubs

Model	DC#	Development Clubs	Obs.	Predicted Per Capita Income
0	1	EXPROP8497 ≤ 8.400, ELF60 > 0.605	31	1616.21
	2	EXPROP8497 ≤ 8.400, ELF60 ≤ 0.605	30	4855.61
	3	EXPROP8497 > 8.40	23	22337.45
1	1	ICRG8497 ≤ 6.520, ELF60 > 0.605	29	1366.98
	2	ICRG8497 ≤ 6.520, ELF60 ≤ 0.605	27	4259.34
	3	6.520 < ICRG8497 ≤ 7.561	9	8671.44
	4	ICRG8497 > 7.561	22	22758.41
2	1	ICRG8497 ≤ 6.520, ELF60 > 0.605	28	1367.15
	2	ICRG8497 ≤ 6.520, ELF60 ≤ 0.605	28	4234.33
	3	ICRG8497 > 6.520	32	17376.26
3	1	ICRG8497 ≤ 6.570, ETHNIC > 0.671	27	1213.81
	2	ICRG8497 ≤ 6.570, ETHNIC ≤ 0.671	32	4129.91
	3	ICRG8497 > 6.570	32	17193.73
4	1	ICRG8497 ≤ 6.570, ETHNIC > 0.671	26	1208.43
	2	ICRG8497 ≤ 6.570, ETHNIC ≤ 0.671	33	4090.33
	3	6.570 < ICRG8497 ≤ 7.561	9	8671.44
	4	ICRG8497 > 7.561	24	22533.09
5	1	EXPROP8497 ≤ 8.400, ELF60 > 0.605	32	1607.60
	2	EXPROP8497 ≤ 8.400, ELF60 ≤ 0.605	32	4732.79
	3	EXPROP8497 > 8.40	23	22337.45
6	1	EXPROP8497 ≤ 8.400, ELF60 > 0.605	31	1616.21
	2	EXPROP8497 ≤ 8.400, ELF60 ≤ 0.605	33	4694.19
	3	EXPROP8497 > 8.40	24	22401.20
7	1	EXPROP8497 ≤ 8.400, ETHNIC > 0.671	28	1295.13
	2	EXPROP8497 ≤ 8.400, ETHNIC ≤ 0.671	42	4693.58
	3	EXPROP8497 > 8.40	27	20338.28
8	1	EXPROP8497 ≤ 8.400, ETHNIC > 0.671	27	1292.72
	2	EXPROP8497 ≤ 8.400, ETHNIC ≤ 0.671	43	4665.08
	3	EXPROP8497 > 8.400	28	20456.38

Table 4: Regression Trees for Multiple Regimes

Model	MR#	Multiple Growth Regimes	Obs.
0	1	EXPROP8497 ≤ 7.525	75
	2	EXPROP8497 > 7.525	72
1	1	ICRG8497 ≤ 7.045	97
	2	ICRG8497 > 7.045	54
2	-	No splits	155
3	1	ICRG8497 ≤ 7.045	100
	2	ICRG8497 >7.045	56
4	1	ICRG8497 ≤ 6.476	93
	2	ICRG8497 >6.476	67
5	-	No splits	151
6	1	EXPROP8497 ≤ 7.525	81
	2	EXPROP8497 > 7.525	74
7	1	EXPROP8497 ≤ 7.525	81
	2	EXPROP8497 > 7.525	75
8	1	EXPROP8497 ≤ 7.525	83
	2	EXPROP8497 > 7.525	77

Table 5[∇]: Development Clubs Country List for Models 0 to 8 (M0-M8)

Country	M0	M1	M2	M3	M4	M5	M6	M7	M8
W. Europe & Offshoots									
Austria	3^+	4	3	3	4	3	3	3	3
Australia	3^+	4	3	3	4	3	3	3	3
Belgium	3^+	4	3	3	4	3	3	3	3
Canada	3^+	4	3	3	4	3	3	3	3
Denmark	3^+	4	3	3	4	3	3	3	3
Finland	3^+	4	3	3	4	3	3	3	3
France	3^+	4	3	3	4	3	3	3	3
Greece	2	3	3	3	3	2	2	2	2
Iceland			3		4		3		3
Ireland	3^+	4	3	3	4	3	3	3	3
Israel	3^+	4	3	3	4	3	3	3	3
Italy	3^+	4	3	3	4	3	3	3	3
Netherlands	3^+	4	3	3	4	3	3	3	3
New Zealand	3^+	4	3	3	4	3	3	3	3
Norway	3^+	4	3	3	4	3	3	3	3
Portugal	3^+	4	3	3	4	3	3	3	3
Spain	3^+	4	3	3		3	3	3	3
Sweden	3^+	4	3	3	4	3	3	3	3
Switzerland	3^+	4	3	3	4	3	3	3	3
United Kingdom	3^+	4	3	3	4	3	3	3	3
United States	3^+	4	3	3	4	3	3	3	3
Asia									
Bangladesh				2	2			2	2
China				2	2			2	2
Hong Kong	3^+	4	3	3	4	3	3	3	3
India	1	1	1	2	2	1	1	2	2
Indonesia	1	1	1	1	1	1	1	1	1
Japan	3^+	4	3	3	4	3	3	3	3
Korea,Republic of	3^+	3	3	3	3	3	3	3	3
Malaysia	1	3	3	3	3	1	1	2	2
Pakistan	1_-	1	1	1	1	1	1	1	1
Papua New Guinea	2	2	2	2	2	2	2	2	2
Philippines	1_-	1	1	2	2	1	1	2	2
Singapore	3^+	4	3	3	4	3	3	3	3
Sri Lanka	2_-	2	2	2	2	2	2	2	2
Taiwan				3	4			3	3
Thailand	1	3	3	3	3	1	1	2	2

[∇] "+" denotes countries above Hansen's 95% CI bound while "-" denotes countries below.

Table 5$^{\nabla}$ (cont.): Development Clubs Country List for Models 0 to 8 (M0-M8)

Country	M0	M1	M2	M3	M4	M5	M6	M7	M8
Latin America & the Caribbean									
Argentina	2_{-}	2	2	2	2	2	2	2	2
Bolivia	1_{-}	1	1	1	1	1	1	1	1
Brazil	2	3	3	3	3	2	2	2	2
Chile	2	3	3	3	3	2	2	2	2
Colombia	2	2	2	2	2	2	2	2	2
Costa Rica	2	3	3	3	3	2	2	2	2
Dominican Republic	2_{-}	2	2	2	2	2	2	2	2
Ecuador	2_{-}	2	2	2	2	2	2	2	2
El Salvador		2	2			2	2		
Guatemala	1_{-}	1	1	2	2	1	1	2	2
Guyana			2		2		2		2
Haiti		2	2			2	2		
Honduras	2_{-}	2	2	2	2	2	2	2	2
Jamaica	2	2	2	2	2	2	2	2	2
Mexico	2	2	2	2	2	2	2	2	2
Nicaragua	2_{-}	2	2	2	2	2	2	2	2
Panama	2_{-}	2	2	2	2	2	2	2	2
Paraguay	2	2	2	2	2	2	2	2	2
Peru	2_{-}	2	2	2	2	2	2	2	2
Trinidad & Tobago	2	2	2	2	2	2	2	2	2
Uruguay	2	2	2	2	2	2	2	2	2
Venezuela	2	2	2	2	2	2	2	2	2
Middle East & North Africa									
Algeria	2_{-}	2	2	2	2	2	2	2	2
Egypt	2	2	2	2	2	2	2	2	2
Iran				2	2			2	2
Jordan	2	2	2	2	2	2	2	2	2
Lebanon				2	2			2	2
Morocco	2	2	2	2	2	2	2	2	2
Syrian Arab Rep.	2_{-}	2	2	2	2	2	2	2	2
Tunisia	2_{-}	2	2	2	2	2	2	2	2
Turkey	2	2	2	2	2	2	2	2	2

$^{\nabla}$ “+” denotes countries above Hansen’s 95% CI bound while “-” denotes countries below.

Table 5[∇] (cont.): Development Clubs Country List for Models 0 to 8 (M0-M8)

Country	M0	M1	M2	M3	M4	M5	M6	M7	M8
Sub-Saharan Africa									
Angola		1		1		1		1	
Botswana	2	3	3	3	3	2	2	2	2
Burkina Faso	1_-	1	1	1	1	1	1	1	1
Cameroon	1_-	1	1	1	1	1	1	1	1
Congo	1_-	1	1	1	1	1	1	1	1
Cote d'Ivoire	1	1	1	1	1	1	1	1	1
Ethiopia	1_-	1	1	1	1	1	1	1	1
Gabon	1	1	1	1	1	1	1	1	1
Gambia	1	1	1	1	1	1	1	1	1
Ghana	1_-	1	1	1	1	1	1	1	1
Guinea	1_-	1	1	1	1	1	1	1	1
Guinea Bissau				1	1			1	1
Kenya	1_-	1	1	1	1	1	1	1	1
Madagascar	2_-	2	2	1	1	2	2	1	1
Malawi	1	1	1	1	1	1	1	1	1
Mali	1_-	1	1	1	1	1	1	1	1
Mozambique	1_-	1	1	1	1	1	1	1	1
Niger	1_-	1	1	2	2	1	1	2	2
Nigeria	1_-	1	1	1	1	1	1	1	1
Senegal	1_-	1	1	1	1	1	1	1	1
Sierra Leone	1_-	1	1	1	1	1	1	1	1
South Africa	1	3	3	3	3	1	1	1	1
Tanzania	1	1	1	1	1	1	1	1	1
Togo	1_-	1	1	1	1	1	1	1	1
Uganda	1_-	1	1	1	1	1	1	1	1
Zaire	1_-	1	1	1	1	1	1	1	1
Zambia	1_-	1	1	1	1	1	1	1	1
Zimbabwe	2_-	2	2	2	2	2	2	2	2

[∇] "+" denotes countries above Hansen's 95% CI bound while "-" denotes countries below.

Table 6[ϒ]: Multiple Regimes Country List for Models 0 to 8 (M0-M8)

Country	M0	M1	M2	M3	M4	M5	M6	M7	M8
W. Europe & Offshoots									
Austria	2^{+}	2	x	2	2	x	2	2	2
Australia	2^{+}	2	x	2	2	x	2	2	2
Belgium	2^{+}	2	x	2	2	x	2	2	2
Canada	2^{+}	2	x	2	2	x	2	2	2
Denmark	2^{+}	2	x	2	2	x	2	2	2
Finland	2^{+}	2	x	2	2	x	2	2	2
France	2^{+}	2	x	2	2	x	2	2	2
Greece	2	2	x	2	2	x	2	2	2
Iceland			x		2		2		2
Ireland	2^{+}	2	x	2	2	x	2	2	2
Israel	2^{+}	2	x	2	2	x	2	2	2
Italy	2^{+}	2	x	2	2	x	2	2	2
Netherlands	2^{+}	2	x	2	2	x	2	2	2
New Zealand	2^{+}	2	x	2	2	x	2	2	2
Norway	2^{+}	2	x	2	2	x	2	2	2
Portugal	2^{+}	2	x	2	2	x	2	2	2
Spain	2^{+}	2	x	2	2	x	2	2	2
Sweden	2^{+}	2	x	2	2	x	2	2	2
Switzerland	2^{+}	2	x	2	2	x	2	2	2
United Kingdom	2^{+}	2	x	2	2	x	2	2	2
United States	2^{+}	2	x	2	2	x	2	2	2
Asia									
Bangladesh				1	1			1	1
China				1	2			2	2
Hong Kong	2^{+}	2	x	2	2	x	2	2	2
India	2^{+}	1	x	1	1	x	2	2	2
Indonesia	2	1	x	1	1	x	2	2	2
Japan	2^{+}	2	x	2	2	x	2	2	2
Korea,Republic of	2^{+}	2	x	2	2	x	2	2	2
Malaysia	2^{+}	2	x	2	2	x	2	2	2
Pakistan	1_{-}	1	x	1	1	x	1	1	1
Papua New Guinea	2	1	x	1	1	x	2	2	2
Philippines	1_{-}	1	x	1	1	x	1	1	1
Singapore	2^{+}	2	x	2	2	x	2	2	2
Sri Lanka	1_{-}	1	x	1	1	x	1	1	1
Taiwan				2	2			2	2
Thailand	2	1	x	1	2	x	2	2	2

[ϒ] Tree regressions for Models 2 and 5 turned up evidence for single regimes ("x" denotes countries for each sample). "+" denotes countries above Hansen's 95% CI bound while "-" denotes countries below.

Table 6[ϒ] (cont.): Multiple Regimes Country List for Models 0 to 8 (M0-M8)

Country	M0	M1	M2	M3	M4	M5	M6	M7	M8
Latin America & the Caribbean									
Argentina	1_{-}	1	x	1	1	x	1	1	1
Bolivia	1_{-}	1	x	1	1	x	1	1	1
Brazil	2^{+}	1	x	1	2	x	2	2	2
Chile	2	1	x	1	2	x	2	2	2
Colombia	2	1	x	1	1	x	2	2	2
Costa Rica	1	1	x	1	2	x	1	1	1
Dominican Republic	1_{-}	1	x	1	1	x	1	1	1
Ecuador	1_{-}	1	x	1	1	x	1	1	1
El Salvador		1	x			x	1		
Guatemala	1_{-}	1	x	1	1	x	1	1	1
Guyana			x		1		1		1
Haiti		1	x			x	1		
Honduras	1_{-}	1	x	1	1	x	1	1	1
Jamaica	1	1	x	1	1	x	1	1	1
Mexico	2	1	x	1	1	x	2	2	2
Nicaragua	1_{-}	1	x	1	1	x	1	1	1
Panama	1_{-}	1	x	1	1	x	1	1	1
Paraguay	1	1	x	1	1	x	1	1	1
Peru	1_{-}	1	x	1	1	x	1	1	1
Trinidad & Tobago	1	1	x	1	1	x	1	1	1
Uruguay	1	1	x	1	1	x	1	1	1
Venezuela	1	1	x	1	1	x	1	1	1
Middle East & North Africa									
Algeria	1_{-}	1	x	1	1	x	1	1	1
Egypt	1	1	x	1	1	x	1	1	1
Iran				1	1			1	1
Jordan	1	1	x	1	1	x	1	1	1
Lebanon									
Morocco									
Syrian Arab Rep.	1_{-}	1	x	1	1	x	1	1	1
Tunisia	1_{-}	1	x	1	1	x	1	1	1
Turkey	2	1	x	1	1	x	2	2	2

[ϒ] Tree regressions for Models 2 and 5 turned up evidence for single regimes. The "x" denotes countries for each sample. "+" denotes countries above Hansen's 95% CI bound while "-" denotes countries below.

Table 6[ϒ] (cont.): Multiple Regimes Country List for Models 0 to 8 (M0-M8)

Country	M0	M1	M2	M3	M4	M5	M6	M7	M8
Sub-Saharan Africa									
Angola									
Botswana	2^{+}	2	x	2	2	x	2	2	2
Burkina Faso									
Cameroon	1_{-}	1	x	1	1	x	1	1	1
Congo									
Cote d'Ivoire									
Ethiopia									
Gabon									
Gambia	2^{+}	1	x	1	1	x	2	2	2
Ghana	1_{-}	1	x	1	1	x	1	1	1
Guinea									
Guinea Bissau				1	1			1	1
Kenya	1_{-}	1	x	1	1	x	1	1	1
Madagascar									
Malawi	1	1	x	1	1	x	1	1	1
Mali	1_{-}	1	x	1	1	x	1	1	1
Mozambique	1_{-}	1	x	1	1	x	1	1	1
Niger	1_{-}	1	x	1	1	x	1	1	1
Nigeria									
Senegal	1_{-}	1	x	1	1	x	1	1	1
Sierra Leone	1_{-}	1	x	1	1	x	1	1	1
South Africa	1	2	x	2	2	x	1	1	1
Tanzania	1	1	x	1	1	x	1	1	1
Togo	1_{-}	1	x	1	1	x	1	1	1
Uganda	1_{-}	1	x	1	1	x	1	1	1
Zaire	1_{-}	1	x	1	1	x	1	1	1
Zambia	1_{-}	1	x	1	1	x	1	1	1
Zimbabwe	1_{-}	1	x	1	1	x	1	1	1

[ϒ] Tree regressions for Models 2 and 5 turned up evidence for single regimes. The "x" denotes countries for each sample. "+" denotes countries above Hansen's 95% CI bound while "-" denotes countries below.

Table 7[+]: OLS and Regression Tree Estimates for Growth Regimes

	MRW	Model 0		Model 1		Model 2	Model 3	
		MR1	MR2	MR1	MR2		MR1	MR2
CONST	-0.6776	-1.3002	0.9351*	-1.6275*	2.6052**	-0.7405	-1.6433**	2.9880**
	(0.4639)	(0.9073)	(0.5104)	(0.8454)	(0.4881)	(0.4878)	(0.8233)	(0.4750)
MNGD	-0.6869**	-0.6785*	-0.3735*	-0.9172**	-0.3530**	-0.5981**	-0.9915**	-0.4072**
	(0.1912)	(0.3588)	(0.2193)	(0.3282)	(0.1704)	(0.2048)	(0.3137)	(0.1767)
MINV	0.3206**	0.2070**	0.5375**	0.2451**	0.4700**	0.3256**	0.2701**	0.4310**
	(0.0398)	(0.0555)	(0.0657)	(0.0498)	(0.0756)	(0.0440)	(0.0482)	(0.0766)
MSCH15	0.0691**	0.0432**	0.0510**	0.0508**	0.0403**	0.0538	0.0625**	0.0523**
	(0.0150)	(0.0219)	(0.0214)	(0.0211)	(0.0173)	(0.1704)	(0.0189)	(0.0173)
DUM6079	0.2957**	0.3517**	0.1307**	0.3302**	-0.0264	0.2917**	0.3313**	-0.0481
	(0.0457)	(0.0729)	(0.0507)	(0.0653)	(0.0452)	(0.0492)	(0.0632)	(0.0463)
MGDP0	-0.2008**	-0.1073*	-0.3480**	-0.1468**	-0.4905**	-0.1713**	-0.1734**	-0.5333**
	(0.0372)	(0.0561)	(0.0503)	(0.0518)	(0.0469)	(0.0412)	(0.0487)	(0.0451)
N	176	75	72	97	54	155	100	56

[+] Dependent variable is the growth rate of real GDP per capita across, respectively, the periods 1960-79 and 1980-99. Standard errors are in parentheses. Model specifications are described in Table 1. "**" indicates significance at the 5% level while "*" indicates significance at the 10% level.

Table 7+ (cont.): OLS and Regression Tree Estimates for Growth Regimes

	Model 4		Model 5	Model 6		Model 7		Model 8	
	MR1	MR2		MR1	MR2	MR1	MR2	MR1	MR2
CONST	-1.1388	1.8841**	-0.8239*	-1.1184	0.9173	-1.2772	1.1826**	-1.0719	1.1624**
	(0.7865)	(0.5202)	(0.4987)	(0.8390)	(0.5010)	(0.8778)	(0.5315)	(0.8041)	(0.5229)
MNGD	-0.6898**	-0.4130**	-0.6672**	-0.6139*	-0.3664	-0.6859**	-0.4119*	-0.5890*	-0.4049*
	(0.2960)	(0.1983)	(0.2140)	(0.3160)	(0.2152)	(0.3424)	(0.2245)	(0.3028)	(0.2208)
MINV	0.2212**	0.4464**	0.3318**	0.1906**	0.5398**	0.2229**	0.5295**	0.2160**	0.5326**
	(0.0494)	(0.0848)	(0.0445)	(0.0540)	(0.0644)	(0.0526)	(0.0693)	(0.0514)	(0.0679)
MSCH15	0.0494**	0.0527**	0.0546**	0.0408*	0.0496**	0.0436**	0.0659**	0.0419**	0.0644**
	(0.0195)	(0.0195)	(0.0172)	(0.0221)	(0.0210)	(0.0200)	(0.0213)	(0.0197)	(0.0207)
DUM6079	0.3234**	0.0500	0.2901**	0.3231**	0.1321**	0.3226**	0.1249**	0.3167**	0.1263**
	(0.0650)	(0.0492)	(0.0498)	(0.0720)	(0.0496)	(0.0695)	(0.0520)	(0.0683)	(0.0510)
MGDP0	-0.1292**	-0.4268**	-0.1850**	-0.1023*	-0.3447**	-0.1148**	-0.3822**	-0.1074**	-0.3790**
	(0.0513)	(0.0427)	(0.0427)	(0.0544)	(0.0486)	(0.0535)	(0.0460)	(0.0522)	(0.0446)
N	93	67	151	81	74	81	75	83	77

+ Dependent variable is the growth rate of real GDP per capita across, respectively, the periods 1960-79 and 1980-99. Standard errors are in parentheses. Model specifications are described in Table 1. "**" indicates significance at the 5% level while "*" indicates significance at the 10% level.