WP/08/263

IMF Working Paper

What is Really Good for Long-Term Growth? Lessons from a Binary Classification Tree (BCT) Approach

Rupa Duttagupta and Montfort Mlachila

INTERNATIONAL MONETARY FUND

 WP/08/263

IMF Working Paper

Western Hemisphere Department

What is Really Good for Long-Term Growth? Lessons from a Binary Classification Tree (BCT) Approach

Prepared by Rupa Duttagupta and Montfort Mlachila[1]

Authorized for distribution by Martín Cerisola

December 2008

Abstract

Although the economic growth literature has come a long way since the Solow-Swan model of the fifties, there is still considerable debate on the "real' or "deep" determinants of growth. This paper revisits the question of what is really important for strong long-term growth by using a Binary Classification Tree approach, a nonparametric statistical technique that is not commonly used in the growth literature. A key strength of the method is that it recognizes that a combination of conditions can be instrumental in leading to a particular outcome, in this case strong growth. The paper finds that strong growth is a result of a complex set of interacting factors, rather than a particular set of variables such as institutions or geography, as is often cited in the literature. In particular, geographical luck and a favorable external environment, combined with trade openness and strong human capital are conducive to growth.

JEL Classification Numbers: O40, 057, O67, C14

Keywords: economic growth, Binary Classification Tree

Author's E-Mail Address: rduttagupta@imf.org, mmlachila@imf.org

[1] We thank Martín Cerisola, Marcos Chamon, Alejandro Guerson, Vladimir Klyuev, Brieuc Monfort, Martin Mülheisen, Steve Phillips, Shaun Roache, Alejandro Santos, Erica Tsounta, and participants at a Western Hemisphere Department seminar for helpful comments and suggestions. The usual disclaimer applies.

Contents Page

Tables

Figures

Appendix

"The number of 'growth regressions' has grown faster than the economies they analyze."
Hendry and Krolzig (2004).

I. Introduction

There is no generally accepted theory of economic growth. The Solow-Swan (1956) neo-classical model is probably the most commonly used basis for running growth regressions. Basically, the model states that in the steady state, growth rates depend on the rates of growth of the labor force and of technological progress. In practice, this is of little value empirically as the actual determinants of technological progress are not specified. As a result, there has been an explosion of the growth literature as noted by Hendry and Krolzig (2004), with each author attempting to claim superiority of their unique dataset, robustness of their instrumental variables, sophistication of econometric technique, or the extent of coverage in the database (number of variables, countries, and periods).

Empirically, two main approaches have been used to analyze growth. The most frequently used approach is the extreme-bounds analysis à la Leamer (1983, 1985). In this approach, a coefficient of interest is deemed robust if it displays a small variation relative to the presence or absence of other regressors, i.e., if its extreme bounds (defined as the coefficient plus/minus two times its standard error) lie to one side or the other of zero. In a bid to put order into the chaos of growth regressions, Levine and Renelt (1992) used extreme-bounds analysis, but unfortunately found that virtually all the variables are "fragile," and that few are robust. In a less restrictive variant of Levine and Renelt (1992), i.e., only if 95 percent of coefficient estimates lie to one side of zero is a variable considered robust, Sala-i-Martin (1997), after running millions of regressions, found that a lot more variables (over 20) are robust.

In a variant of the above approach, Sala-i-Martin, Doppelhoffer and Miller (2004, hereafter "SDM") use a Bayesian averaging of classical estimates (BACE) approach to examine the robustness of explanatory variables in cross-country growth regressions. They argue that their method is superior to the traditional approach as it requires only a guess on the size of one hyper-parameter, the expected size of the model, $\bar{k}$. Moreover, their parameter estimates can be calculated using repeated applications of ordinary least squares (OLS). They find that out of 67 variables, 21 variables proxying for a wide range of economic, geographic, and structural indicators are significantly correlated with long-term growth

The other approach to modeling growth is the so-called LSE approach, or the general-to-specific modeling. The proponents of this approach, such as Hoover and Perez (2004) and Hendry and Krolzig (2004), claim that their "method has proved superior to extreme-bounds methods in isolating the truth—when the truth is to be found" (Hoover and Perez 2004). They claim that their models: (i) are statistically well-specified, i.e., with white noise errors; (ii) are valid restrictions of the general model; and (iii) encompass every other parsimonious regression that is a valid restriction of the general model; and (iv) are very cost effective ("We ran one equation", Hendry and Krolzig 2004). In contrast to the extreme bounds approach, Hoover and Perez (2004) and Hendry and Krolzig (2004) find that there are in fact only five variables that can robustly explain growth, in ascending order: (i) number of coups and revolutions; (ii) fraction of the population that is protestant; (iii) fraction of the population that is Confucian; (iv) equipment investment as a fraction of GDP; and (v) the number of years a country has been an open economy.

Interestingly, Hoover and Perez (2004) give short shrift to most of the favorite growth theories, such as "institutions rule," "it's geography, stupid', financial factors, health or human development. As underscored by SDM, "the multiplicity of possible regressors is one of the major difficulties faced by researchers trying to make sense of the empirical evidence on economic growth." They argue that this typically leads to "creative theorizing" and the worst forms of data mining that can lead to spurious conclusions.

As can be seen, two of the studies that each claims to be the "gold standard" (Hoover and Perez 2004 and SDM 2004) only have two common significant variables—fraction of population that is Confucian and the number of years an economy is open—out of the universe of possible variables. These two common variables do not give much comfort as to what is really important for growth. What to make of this? What are the relative roles of structural characteristics vis-à-vis policies in influencing growth? Are there threshold effects? Do countries with strong growth have something special?

Our paper aims to contribute to the literature by offering some answers to the above questions using the binary classification tree (BCT), a nonparametric statistical technique. A key strength of the method is that it recognizes that a combination of conditions can be instrumental in leading to a particular outcome, thereby making room for non-linearities in variable interactions. Moreover, the BCT approach does not prescribe a particular functional form—not a trivial decision in growth regressions. Unlike the general approach in the literature, the approach used here can therefore distinguish more easily the factors that determine growth in low-, mid-, and high-growth countries, without an *a priori* theoretical construct. In this paper, we mainly focus on key factors that are particularly good for a strong growth performance. In alternative specifications that focus exclusively on indicators that have some policy content, the paper also provides insight into the "do's" and "don'ts" of growth by identifying policies important for strong growth as well those underlying poor growth.

To do this we use the SDM database which contains some of the most frequently used variables in the growth literature when accounting for both policy indicators and structural characteristics. We therefore do not construct our own variables, but effect a few updates. We subject the various variables to the BCT to determine the most important variables. Our study finds that growth is a result of a complex set of interacting factors; in particular, geographical luck and a favorable external environment, combined with strong economic policies that promote greater trade openness and the development of human capital are really good for growth. Focusing only on policy variables, we confirm that policies geared toward trade openness, combined with strong human capital indicators, such as education and relatively modest population growth are conducive to growth. Conversely, poor human capital indicators combined with trade distortions could lead to a poor growth outcome. That said, we do not attempt to develop a new growth theory in the sense of analyzing the deep reasons why particular combinations of variables seem to matter.

The rest of the paper is organized as follows. The next section presents a quick overview of the development of the growth literature in order to highlight some of the key debates and pitfalls. Section III describes the BCT methodology with a particular emphasis on its strengths relative to more traditional approaches. Section IV describes the dataset and highlights a few stylized facts. Section V presents the principal empirical results and robustness tests. Section VI provides some concluding remarks.

II. A Few Notes on the Growth Literature

This section provides a quick overview of the main issues tackled by the economic growth literature over the past two decades. It is not meant to be a comprehensive review;[2] rather, the main point is to highlight the main problems identified by the literature and how they have been typically tackled. It focuses on a few papers published over the past few years, notably SDM, Hover and Perez (2004), and Rodrik *et al.* (2004).

The growth of the growth literature over the past two decades is quite phenomenal, starting with Romer (1986), among others.[3] The revival of the so-called new growth literature started with a critique of the Solow-Swan framework in which the rate of growth is exogenously determined as it is a function of the growth of the labor force and technological progress. The main problem with this approach is that the theory does not explain the factors that determine labor force growth and technological progress.

The subsequent development of the endogenous growth theory was a response to these shortcomings. This theory focuses on the micro-foundations that determine the production of human capital and technological progress. The theory also shows how policies can have an impact on these two factors, for instance, through expenditure on education, and research and development.

Several problems have been identified by the empirical literature in trying to implement this theoretical approach. The main problem, as highlighted in the introduction is that there is little guidance as to which factors affect long-term growth and how. As noted by SDM, there is a very large number of potential regressors which could exceed the number of countries, thus leading to computational impossibilities. This "open-endedness of theories" (or model

[2] Barro and Sala-i-martin (2004) is an excellent general reference work on growth and offers a comprehensive bibliography.

[3] It has spawned its own journal, the *Journal of Economic Growth*. As potential areas of inquiry fit for publication in the journal, it lists no less seven potential groups of variables related to growth, including, financial development, human capital, income distribution, fertility, politics, and trade.

uncertainty), as Brock and Durlauf (2000) called it, poses a serious epistemological problem as the "validity of one causal theory does not imply the falsity of another."

To overcome the multiplicity of regressors, the literature has adopted various approaches. We highlight only two of them: the Bayesian and the general-to-specific modeling approaches. SDM is typical of the Bayesian approach. Their basic objective is to overcome the worst excesses of data mining that often leads to creative theorizing by reducing the number of parameters that have to be considered when setting up an empirical framework. In their case, using a balanced cross-section sample of 67 variables over a period of four decades, they find that there are 18 variables that are robustly linked to growth. The list of robust variables that they identify includes almost the full spectrum that other authors have found to be significant, including human capital (such as primary school education and life expectancy), geography (e.g., malaria prevalence and tropical areas), conditional convergence of growth (initial income level), role of raw material exports (such as mining), cultural and religious influences (e.g., Buddhism), and macroeconomic ones (e.g., the relative price of investment openness, share of government consumption in GDP, etc.).

Another influential approach is the use of general-to-specific modeling developed and promoted by Hendry, among others. This approach claims that it is superior to other methods such as the extreme bounds approach when it comes to isolating the "truth"—if it is present—from a set of variables. The approach allows one to choose a dominant and encompassing model to overcome the issue of model uncertainty. The proponents of the approach claim that it allows for parsimonious and statistically well-specified models with valid restrictions of the general regression. That said, the approach does not lead to what could be called unified theory of growth, as it is often dependent on the underlying dataset. While for a given dataset the approach will typically find a single well-specified model, different datasets can indeed turn up with different models. As an example, Hoover and Perez (2004), and Bleaney and Nishiyama (2002), find different variables to be key determinants of growth.

The second problem that the literature has had to grapple with—especially the extreme bounds approach—is endogeneity. As is well-known, the presence of endogeneity in a regression nullifies inferences on causal relationships. The literature has typically dealt with this issue by developing instrumental variables. For instance, Acemoglu *et al.* (2004) provide an instrumental variable that has proven quite popular—the mortality rate of colonial settlers. This is typically used as an instrument for institutional quality. Rodrik *et al*. (2004) use this instrument to determine which are the "deep" determinants of economic growth.[4] They argue that institutions trump everything including the two main factors often cited in the literature—geography (as in Sachs 2001) and international trade (as in Sachs and Warner 1995).

[4] Strictly speaking, they do not model economic growth but rather determinants of economic development as measured by the level of PPP per capita GDP in 1995.

Table 1. Most Significant Variables in Selected Growth Studies

Paper	Sala-i-Martin et al. (2004)	Rodrik et al. (2004)	Hoover and Perez (2004)	Bleaney and Nishiyama (2002)
Methodology	Bayesian averaging of classical estimates	Panel regression with horse-race between institutions, geography and integration using instrumental variables	General-to-specific parsimonious model using Sala-i-Martin (1997) data	Benchmark encompassing model of Barrro (1997), Easterly and Levine (1997) and Sachs and Warner (1997) using general-to-specific modeling
Dependent variable	Average per capita GDP growth rate 1960 to 1996 (PPP adjusted)	Log of GDP per capita in 1995 (PPP adjusted). Also income per worker, capital per worker, human capital per work, and total factor productivity		Average per capita GDP growth rate 1965 to 1990 (PPP adjusted)
Main variables	East Asian dummy Primary schooling in 1960 Investment price GDP in 1960 (log) Fraction of tropical area Coastal population density (1960s) Malaria prevalence (1960s) Life expectancy 1960 Fraction Confucian African dummy Latin American dummy Fraction of mining in GDP Spanish colony Years economy is open Fraction of population Muslim Fraction of population Buddhist Ethnolinguistic fractionalization Share of govt. in consumption (1960s)	Rule of law	Coups and revolutions Fraction Confucian Equipment investment/GDP Fraction protestant Years economy is open	Log of per capita GDP in 1965 (Y) Y^2 Openness Openness x Y Log of life expectancy in 1965 Male schooling Institutional quality Democracy index Square of democracy index Cent. government consumption /GDP Primary product exports/GDP Terms of trade growth Tropical climate Economically active minus total pop. growth
Other important variables	Population density in 1960 Real exchange rate distortions Fraction of pop. speaking foreign language			

Sources: Sala-i-Martin et al. (2004), Rodrik et al. (2004), Hoover and Perez (2004), and Bleaney and Nishiyama (2002).

A final problem worth mentioning—but for which there are no easy answers—is parameter heterogeneity. Brock and Durlauf (2000) argue that the assumption that parameters are identical across countries—a fundamental logical conclusion of any econometric estimation—is “very implausible especially when studying such complex and heterogeneous objects as countries.” Although it is inevitable in econometrics to find a fairly unique set of parameters, they argue that it is often misleading given pervasive model uncertainty. While Brock and Durlauf probably overstate the extent of the problem given that parameters only provide central tendencies within a range of standard errors, the inherent mixing of countries is problematic, despite the use of regional dummies to overcome it.

In conclusion, while there have been substantial theoretical and especially empirical advances to the growth literature since the mid-eighties, starting with endogenous growth models, the fact remains that there are still raging debates as to the theoretical approach, methodology and variables. Table 1 summarizes the somewhat conflicting evidence from the key studies cited above.

III. The Binary Classification Tree (BCT) Approach[5]

The BCT is a nonparametric statistical technique that is able to sift large databases of variables and identify significant patterns among them to help predict binary outcomes. In this paper, the binary variable takes the value of one for all strong growth performers (as defined below), and zero otherwise. Starting with the whole sample or “parent” node, the BCT compares all candidate variables at all possible threshold values and selects an indicator (and a particular threshold) as a primary splitter based on its ability to split the sample into “purer” sub-samples (or more homogeneous “child” nodes) where the relative proportion of the strong growth outcome either increases or declines significantly compared with the sample average. Thus, at the “parent” node, comprising the entire sample, a primary splitter and its threshold is identified, at which the sample gets split into two child nodes. The process repeats itself at each child node until further splitting is stopped or is impossible. The latter happens when all the cases in the particular node are of the same outcome or there is only one case in the node. In general, however, the tree size is determined in terms of the trade-off between the cost of growing (proportional to the number of nodes in the tree), and the gain from further growing in terms of reducing the percentage of misspecification (i.e., number of outcomes that are erroneously classified, akin to type I and type II errors). When the former offsets the later, the tree stops growing. Finally, each terminal node provides a set of rules that provide the sequence of relationships important for a strong growth outcome (or conversely, the set of rules that relate to a missed growth opportunity).

While the tree shows the splitters and the relationships between them that lead to a particular growth outcome, other important information is provided by the ranking of the candidate

[5] This section largely draws on Duttagupta and Cashin (2008), and the references therein.

variables in terms of their explanatory power for the whole sample. The model computes a score for each variable in terms of its relative importance—in distinguishing a strong growth outcome—at each node of the tree. It is possible for a variable to be slightly outperformed by another as a splitter and therefore never appear in the tree, despite the fact that the variable is better than others in its overall ability to explain strong growth. Hence, while the tree provides important information about the relationship between variables in resulting in a strong growth outcome, it is also important to assess the overall ranking of the available indicators in their ability to explain a strong growth outcome.

The BCT model is particularly suitable for shedding light on a number of unresolved issues in the empirical growth literature.[6] First, given that the literature has not agreed on any specific indicators as being unambiguously important for growth, it is useful to assess whether the ranking of indicators is markedly different using a non-parametric statistical technique such as BCT compared to one that is based on a formal functional form. The most significant contribution from the BCT model, however, comes from the fact that it identifies threshold effects, i.e., it identifies the pattern of relationships between indicators that underscore a strong growth outcome, which would be very difficult to pin down with regression analysis.

In economics, the BCT has been used to analyze various issues of interest, including recent analyses of currency, sovereign debt, capital account and banking crises. To our knowledge, the work by Ghosh and others (1998) is the only other use of BCT to analyze economic growth.[7] The authors consider a panel of 107 countries during 1960-96, and analyze the relationship between growth and 13 conditioning variables. They focus on assessing the key triggers of very strong growth relative to poor growth, by comparing the growth outcomes of the top 33rd percentile of the sample with the bottom 33rd percentile, and find that high investment ratios, strong human capital and low inflation are critical to growth. While the authors use a panel dataset and are hence able to control for cross-country heterogeneity, their database of explanatory variables is relatively narrow, not covering any institutional, religious, and geographical variables. In comparison we use the SDM database, which allows us to consider among a rich spectrum of both policy, and institutional, geographical and structural variables. At the same time, we are able to strip out the structural variables from the database and focus on the policy indicators only to compare our results with those of Ghosh and others (1998).

---

[6] See Duttagupta and Cashin (2008) for further details on the specific advantages and drawbacks of BCT.

[7] There are two related studies: Ghosh and Wolf (1998) and Ghosh and Philips (1998). Wolf (1994) uses the BCT to analyze ranking of variables that explain consumption volatility.

IV. Properties of the Data

The original SDM database is a cross-section of 139 countries with 67 independent variables over the period 1960-96. The dependent variable is the average real per capita GDP growth rate for the sample period. In their regressions, SDM use a balanced data set, i.e., it has an equal number of observations for all regressions. As a result, they use only 88 countries in their analysis. For our approach, since it is not necessary to have a balanced data set, we use the complete set cross-section of 139 countries.[8]

The database of 67 right-hand-side variables that will be used to analyze growth can be separated into six broad categories (Table 2)—economic and human capital policy indicators (29), political policies (7), structural characteristics that proxy for religious, cultural or political orientation (16), and geography (16).[9] Of these, policymakers generally have little or no control over the indicators in the last two broad categories, but have some or full control over the other indicators.

Table 2. Definition of Variables

Economic, and Human Capital Indicators (29)	Political Policies (7)	Strucutural Characterisitics (16)	Geographic Characteristics (16)
Macroeconomic Stability:	(1) Civil liberties	Cultural:	Geographical bearing:
(1) Inflation	(2) Political rights	(1) Ethnological fractionalization	(1) Absolute latitute
(2) Square of inflation	(3) Revolution and coups	(2) Fraction speaking English	(2) East Asian country (dummy)
	(4) Fraction spent in war 1960-90	(3) Fraction speaking other language	(3) European country (dummy)
External policies:	(5) War participation 1960-90		(4) African country (dummy)
(1) Openness measure, 1965-74	(6) Degree of capitalism	Political history:	(5) Latin American country (dummy)
(2) Trade distortion index	(7) Socialist country (dummy)	(1) Colony (dummy)	(6) Landlocked country (dummy)
(3) Outward orientation		(2) British colony (dummy)	(7) Tropical climate zone
(4) Proportion of years open		(3) Spanish colony (dummy)	(8) Fraction of tropical area
		(4) Timing of independence	
External shocks:			Population density:
(1) Terms of trade growth		Religious orientation:	(1) Total population density
(2) Terms of trade ranking		(1) Fraction Buddhist	(2) Population density in the coasts
		(2) Fraction Catholic	(3) Interior population density
Human capital development:		(3) Fraction Orthdox	(4) Fraction of population living in tropics
(1) Fertility rate		(4) Fraction Protestant	
(2) Primary education enrollment		(5) Fraction Confucian	Other:
(3) Higher education enrollment		(6) Fraction Hindu	(1) Air distance between big cities
(4) Life expectancy		(7) Fraction Muslim	(2) Landarea
(5) Malaria prevalence rate		(8) Religious intensity	(3) Hydrocarbon Deposits in 1993
(6) Population growth 1960-90			(4) Closeness to navigable Water
Demographic composition:			
(1) Population in 1960			
(2) Population under 15			
(3) Population over 65			
Government indicators:			
(1) Share of defense spending in GDP			
(2) Share of public education spending in GDP			
(3) Share of public investment in GDP			
(4) Share of government in nominal GDP			
(5) Share of government in real GDP			
(6) Share of government in consumption			
Other indicators			
(1) Initial level of per-capita income			
(2) Size of the economy			
(3) Price of investment			
(4) Share of GDP in mining			
(5) Oil producing country (dummy)			
(6) Share of primary exports in total exports			

Source: Sala-i-Martin et al. (2004).

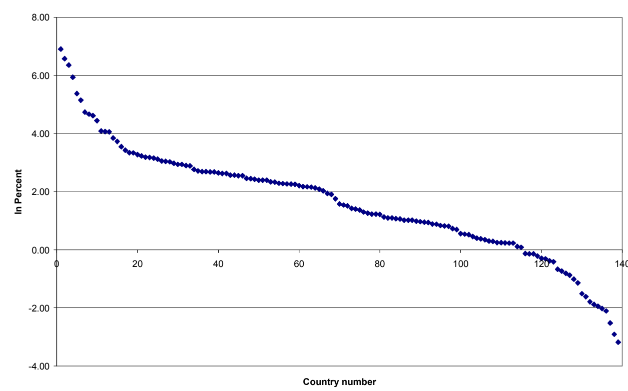
[8] For 16 countries in the SDM sample that have missing growth data, we compute these using data from the World Bank's *World Development Indicators* database These countries are Bahamas, Bahrain, Dominica, Grenada, Hungary, Kuwait, Oman, Poland, Samoa, Solomon Islands, St. Lucia, St. Vincent and the Grenadines, Tonga, United Arab Emirates, Vanuatu, and Yemen.

[9] See Sala-i-Martin and others (2004), and Appendix 1 for a full description of the dataset.

A closer look at the data reveals explanatory variables that could be strongly correlated with each other. For instance, there are four proxies for trade openness—a general openness measure, [10] an indicator which measures the proportion of the years within the sample period that a country practiced openness, an indicator for trade distortion which captures the degree to which relative price differences between countries are explained by trade or real exchange rate distortions, an indicator of outward orientation that combines the third measure with variability in relative price levels.[11] Similarly, there are six indicators proxying for the relative importance of the role of government, four indicators each of population characteristics and population density, eight of geographical bearing, and so on. As a caveat, while this is a very rich dataset—comprising both policy and structural variables—it does not include indicators of governance and financial sector development that could also be potentially important for explaining strong growth.

There is a very wide range of growth rates among the sample countries (Figure 1). Although the sample average for growth is 1.7 percent, the standard deviation is actually higher at 1.9 percent. The top 25 percentile growth performers (our chosen dependent variable) have a cut off at 2.7 percent (with the average growth rate for this group at 3.9 percent), while the cut off for the bottom 25 percentile growth performers is 0.4 percent (and average growth for this group is -0.7 percent). Although the sample is not particularly skewed (skewness of 0.1), it is has a very flat distribution with long tails (kurtosis of - 9.2). The top growth performers are dominated by eastern Asian countries (especially the four tigers—Hong Kong SAR, Korea, Singapore, and Taiwan POC, see Table 3), while the worst performers are African countries, especially those that have undergone war.

Figure 1. Distribution of Growth



V. The Results

A. Baseline Model: What is Good for Strong Growth?

Under the baseline, the dependent variable is a binary variable that takes the value one if average growth in a country during the sample period was in the top 25 percentile of the sample, and zero otherwise. Based on this definition, the unconditional probability of having

---

[10] A country is considered to be open if none of the following five conditions hold: (i) non-tariff barriers covering more than 40 percent of trade; (ii) average tariff rate 40 percent or more; (iii) a black market exchange rate that is 20 percent or more depreciated compared to the official rate; (iv) a socialist economic system; and (v) a state monopoly on major exports. For details, see Sachs and Warner (1995).

[11] For details on the trade distortion and variability measures see, Dollar (1992).

a strong growth outcome is 25 percent, comprising 35 out of a sample of 139 countries. We consider all 66 variables from SDM to identify:[12] (i) which variables rank high in their ability to explain growth; (ii) which indicators are weakly correlated with growth; and (iii) the sequence of relationships between the variables in (i) that underlie a strong growth outcome.

Table 3. Growth Rate for Top Quartile
(in percent)

Country	Growth rate
Singapore	6.9
Taiwan, POC	6.6
Korea	6.4
Hong Kong, SAR	5.9
Malta	5.4
Thailand	5.2
Botswana	4.7
Japan	4.7
Cyprus	4.6
Indonesia	4.5
Portugal	4.1
China	4.1
Malaysia	4.1
St.Lucia	3.8
Ireland	3.7
Spain	3.6
Greece	3.4
Solomon Islands	3.3
St.Vincent & Grens.	3.3
Tunisia	3.3
Seychelles	3.2
Mauritius	3.2
Norway	3.2
Grenada	3.2
Morocco	3.1
Luxembourg	3.1
Cape verde	3.0
Israel	3.0
Italy	3.0
Brazil	2.9
Dominica	2.9
Syria	2.9
Austria	2.9
Gabon	2.8
Finland	2.7
Simple Average	*3.9*
Overall sample	
Average	*1.7*
Median	*1.5*
Standard Deviation	*1.9*
Skewness	*0.1*
Kurtosis	*-9.2*

Source: Sala-i-Martin *et al.* (2004) and authors' calculations.

[12] Note, that although SDM has 67 variables, one of these is the square of inflation, which we exclude from the BCT. In the presence of inflation, the square of inflation does not have any additional explanatory power on growth (alternative monotonic transformations of a given variable are treated alike under the BCT).

Variables strongly related to strong growth performance

Table 4 shows the ranking of the 66 variables in their ability to explain growth, and compares this with SDM's original ranking using the Bayesian methodology.[13] As discussed in Section IV, a particular variable that does not appear in a tree because it is outweighed by another variable at a particular node might still have a high overall ranking owing to the total score based on its relative importance at every node. A few regularities are established:

- Eighteen indicators (out of 66), proxying for a wide spectrum of policies and structural characteristics, meaningfully explain a strong growth outcome: trade and/or real exchange rate distortion, coastal population density, air distance to big cities, overall population density, fraction of population living in the tropics, initial per capita GDP level, school enrollment, terms of trade growth, ethnological fractionalization, share of public investment to GDP, African dummy, life expectancy, political rights, price of investment, civil liberties, absolute latitude, fraction of population under 15, and a measure of religious intensity.

- The results show that while human capital development (school enrollment, life expectancy) is important for growth, so are broad range of economic policy variables, such as the extent of trade distortions, the price of investment, and the level of public investment. These results broadly support the findings of SDM. Moreover, as is common in the literature, the initial level of GDP is also an important predictor for strong growth, as are some exogenous factors such terms of trade growth. Also, seven indicators of geographical and demographic orientation have an important bearing on growth: total population density and that in coastal area, the fraction of population living in tropics—as in SDM—and that under the age 15, the distance between big cities, and the absolute latitude (the latter are not found to be significant by other studies), and an African dummy.

- Other than the African dummy, other dummy variables such as East Asian and Latin American dummies, and a dummy for Spanish colonies are not significantly related to growth, in contrast to the findings of SDM. Also, with the exception of the degree of ethnological fractionalization and religious intensity, other indicators of religious or cultural orientation such as fraction of the population practicing a particular religion are insignificant (many of these are significant in SDM). Finally, most political indicators, with the exception of the degree of civil liberty and political rights are found to be unimportant (as in SDM, but differing from Hoover and Perez, and to some extent BN (who found an index of democracy to be important for growth).

[13] While SDM provides the ranking based on a robust partial correlation of an indicator with growth, ranking in the BCT is made on the basis of the relative importance of the indicator at every node of the tree.

Table 4: What is Really Good for Growth: Ranking of Indicators

	Sala-i-Martin and others (2004)	Duttagupta and Mlachila (2008) Specification (i) Baseline model, with all explanatory variables	Duttagupta and Mlachila (2008) Specification (ii) Excluding G7 countries 1/	Duttagupta and Mlachila (2008) Specification (iii) Excluding HIPC countries 2/
		Variables related to growth		
1	East Asian dummy	Trade distortions	Trade distortions	Population density coastal in 1960s
2	Primary schooling in 1960	Population density coastal in 1960s	Population density coastal in 1960s	Population density 1960
3	Investment price	Air distance to big cities	Air distance to big cities	Trade distortions
4	GDP in 1960 (log)	Population density 1960	Population density 1960	Ethnolinguistic Fractionalization
5	Fraction of tropical area	Fraction population in tropics	Fraction population in tropics	Air distance to big cities
6	Population density coastal	GDP in 1960 (log)	Terms of trade growth in 1960s	Interior population density
7	Malaria prevalence	Primary schooling in 1960	Primary schooling in 1960	Population growth
8	Life expectancy in 1960	Terms of trade growth in 1960s	Ethnolinguistic fractionalization	Terms of trade growth in 1960s
9	Fraction Confucian	Ethnolinguistic fractionalization	Interior population density	English speaking population
10	African dummy	Public investment share	Public investment share	Fraction population in tropics
11	Latin American dummy	African dummy	African dummy	GDP in 1960 (log)
12	Fraction GDP in mining	Life expectancy in 1960	Life expectancy in 1960	Government share of GDP
13	Spanish colony	Political rights	Political rights	Public investment share
14	Years open 1950-94	Investment price	Investment price	Life expectancy in 1960
15	Fraction Muslim	Civil liberties	Public education spending share in GDP	Fraction population less than 15
16	Fraction Buddhist	Absolute latitude	Fraction population less than 15	Public education spending share in GDP
17	Ethnolinguistic fractionalization	Fraction population less than 15	Fraction Hindus	Political rights
18	Gov. consumption share 1960s	Religion measure	GDP in 1960 (log)	Fraction Protestants
19	Population density 1960		Share of government in nominal GDP	Religion measure
20	Trade distortions			Share of government in nominal GDP
21	Fraction speaking foreign language			

1/ The G7 countries are Canada, France, Germany, Italy, Japan, the United Kingdom and the United States
2/ The Highly indebted poor countries in the sample are: Benin, Bolivia, Burkina Faso, Burundi, Cameroon, Chad, Congo, Ethiopia, Gambia, Ghana, Guinea, Guniea Bissau, Guyana, Haiti, Honduras, Malawi, Mali, Mauritania, Mozambique, Nicaragua, Niger, Rwanda, Senegal, Tanzania, Uganda and Zambia.

- Curiously enough, strong growth is not significantly associated with (low) inflation in contrast to the findings of Ghosh and Philips (1998), and Ghosh and Wolf (1998).[14] One possible explanation is that inflation was not a macroeconomic problem in the 1960s or that the growth experience among some of the top growth performers was inflationary. Indeed, a closer look at the data reveals that the average inflation experience of the top 25 percentile growth performers during the 1960s was around 11 percent while that for the rest of the sample countries was around 6 percent. Excluding all countries with less than 50 percent inflation gives an average inflation of the top growth performers similar to that for the rest of the sample at a little above 5 percent.[15] However, the results are broadly supportive of the other findings of Ghosh et. al in that investment ratios, human capital indicators, and initial level of income are significantly related to strong growth.

To summarize, our findings are closer to SDM than to other studies in that like SDM, we do find evidence for a wide spectrum of indicators—ranging from economic and human capital development policies, geographical bearing, and conditional convergence—as having a significant bearing on growth. We also find that a large number of variables are unimportant in explaining growth (48), of which a majority was also found to be insignificant by SDM (37). However, unlike SDM, we do not find strong evidence of geographical dummies being related to growth, with the exception of the African dummy.

Conditions Underlying Strong Growth

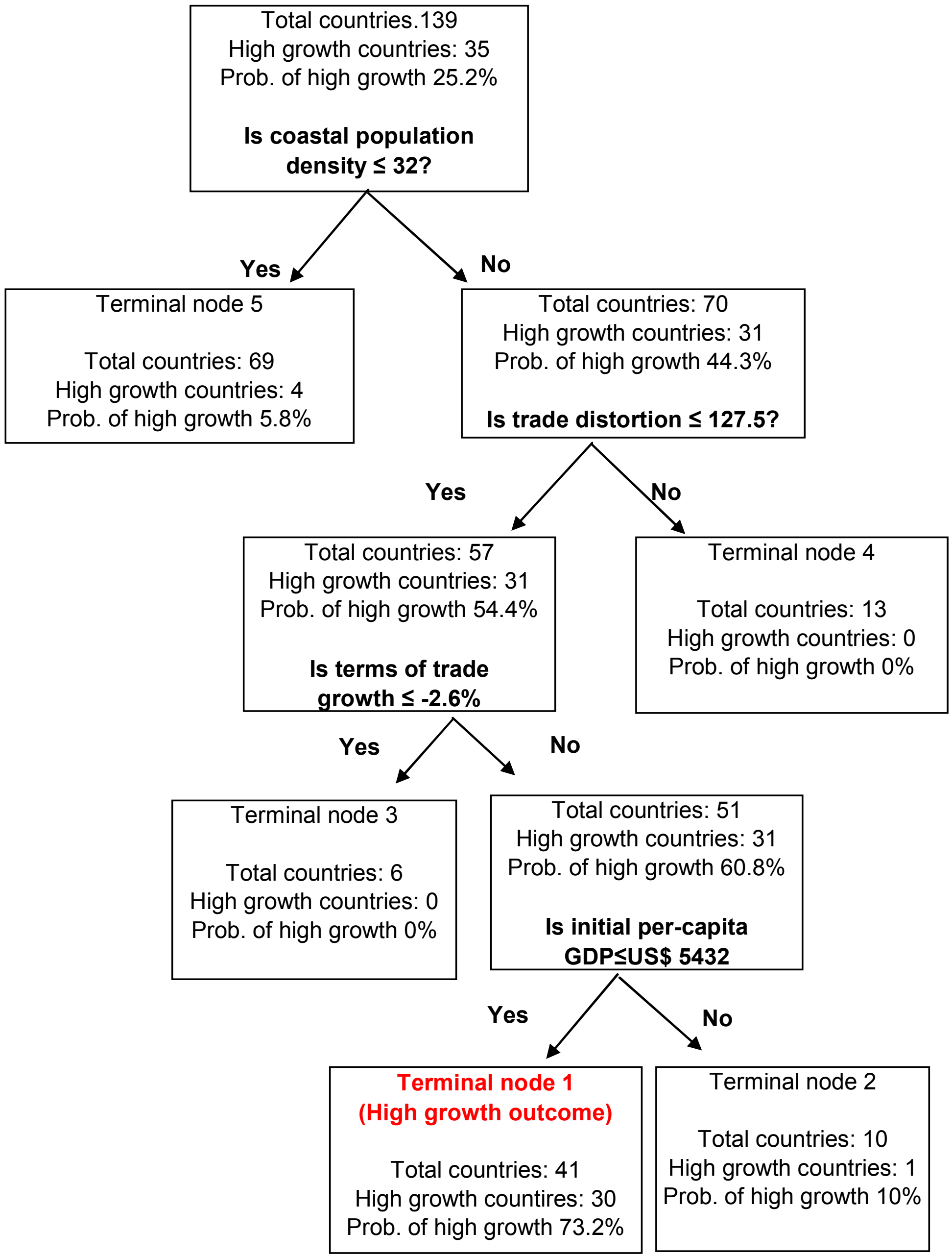
Figure 2 shows the Binary Classification Tree (BCT) under the baseline specification. The tree explores the question on what sequence of relationships between available indicators underlies strong growth. The model has a reasonably good in-sample fit, with 86 percent of the strong growth cases correctly classified, while the type II error (i.e., classifying countries as strong growth performers when they are not) is relatively low at 11 percent. The parent node is split on the basis of population density in coastal areas. As in SDM, a higher coastal population density increases the probability of a strong growth outcome—in particular, for countries with a coastal population density less than 32 people/km^2, the probability of strong growth falls from 25 percent to about 6 percent. Conversely, the probability of a strong growth outcome increases to 44 percent if coastal density is higher than the identified threshold and this child node is split further on the basis of the level of trade distortion. A higher coastal population density could imply first, presence of coasts (which would benefit

[14] The original inflation data in the SDM dataset corresponded to average inflation in 1960-90, raising the issue of endogeneity whereby growth could influence inflation. To get around this, we substitute the inflation data in SDM with average inflation in the 1960s. The results, in terms of the insignificant correlation between growth and inflation, and ranking of the key variables, are unaffected by the choice of the inflation indicator.

[15] Even when the extreme inflation cases are removed from the sample, the BCT does not yield any significant relationship between growth and inflation, and the other key results are unaffected.

trade and business), and also higher population density in a growth-inducing area. The following important relationships are identified at the terminal nodes:

Figure 2. Baseline Model



Conditions underlying strong growth. Terminal node 1 is the only node that is characterized by a strong growth outcome. The combination of factors that increases the conditional probability of a strong growth outcome from 25 percent (in the parent node) to 73 percent are:[16] a coastal population density higher than 32 people/km^2, a trade policy distortions index of less than 127.5 (where 100 implies no distortion), terms of trade growth higher than -2.6 percent, and the initial real per capita GDP less than US$ 5,432. Table 5 shows the median values of the top ten indicators, including the four key splitters in the tree for the full sample, as well as for the strong growth node. As expected, the median values for trade distortion is lower for the strong growth performance node than for the full sample, while that for terms of trade growth and coastal population density is higher. However, the median value of initial income for the strong growth outcome is higher than that for the full sample, although both are much lower than the threshold identified for conditional convergence.

Table 5. Median Values of Key Indicators in Baseline Model

Variables	Ranking	Range 1/	Median Value in Overall Sample	Median Value in Strong Growth Outcome
Trade distortion	1	50-277	116	100
Coastal population density 2/	2	0-3729	30	99
Air distance to big cities 3/	3	140-9,590	4,600	2,558
Total population density	4	0-4179	50	16
Fraction of population living in tropics	5	0-100	5.4	0.0
Initial level of per-capita GDP (US$)	6	257-9,895	1,404	1,648
Primary school enrollment rate	7	0-100	80.0	97.0
Terms of trade growth	8	-7-15	-0.2	0.1
Ethnological fractionalization	9	0-100	27.8	23.8
Share of public investment in GDP	10	0-25	2.3	2.5

Source: Authors' calculations.
1/ Variable min-max values for the full sample.
2/ Number living near coast per unit of area, zero for countries with no coasts.
3/ Logarithm of minimal distance from New York, Rotterdam, or Tokyo.

Conditions representing a missed growth opportunity. Terminal nodes 3, 4, 5 show that relatively low coastal population density (less than 32 people/km^2), or excessive trade distortion (with the index higher than 127), or poor terms of trade growth (less than -2.6 percent) can result in missing a strong growth outcome.

Besides identifying a relatively wide spectrum of economic indicators that matter for growth, the analysis also provides the sequence of relationships between some of the key indicators that could improve the chances of strong growth. In particular:

[16] In the child nodes (or sub-samples) the proportion of observations for each outcome (strong growth or not) gives the conditional probability of that outcome, since it is conditional on meeting a criterion that was used to split the preceding (parent or child) node into subsequent child nodes. Also, at each terminal node, the probability of an outcome is conditional on meeting a sequence of criteria in preceding child-nodes. For example, in the above tree, the probability of having a strong growth outcome in either of the first-tier child nodes (immediately following the parent node) is conditional on the value of population density in the coastal area of the country relative to a threshold of 32 percent.

- The results demonstrate the importance of external sector performance and competitiveness—given by a relatively high population density on the coasts (which are presumably a country's trade hubs), combined with a reasonable external terms of trade growth and relatively low trade distortions—as conducive to growth. This is a significant break from the literature, which has not always pinned down the exact relationship between growth and economic policies.

- Second, the results stress the importance of threshold effects. While high coastal density, low trade distortions, and reasonable terms of trade growth are all good for growth, they cannot individually deliver strong growth independent of the others. In other words, an economy increases its chances of achieving strong growth when it has a relatively high population density in its coasts, *and,* the trade distortion is relatively low, *and* terms of trade growth is reasonable. Under these conditions, the conditional probability of a reasonably strong growth outcome increases from 25 percent (sample average) to 61 percent.

- Finally, the results make a very strong case for the theory of conditional income convergence. For instance, even if all of the above policies related to increased globalization and competitiveness were effective, the chances of a strong growth outcome would still be conditional on initial level of income—the probability of a strong growth outcome increases from 61 percent to 73 percent for countries with initial per capita GDP level less than US$5,432, but declines to 10 percent if initial income is higher than the identified threshold.

B. Alternative Specifications and Robustness Checks

We consider four alternative specifications to check the robustness of the baseline model. Under the first two specifications, we exclude a particular group of countries from the original sample and run the BCT model with the altered smaller sample to assess whether the variable ranking of the baseline model is maintained under the alternative specification, and also to check the "out of sample" forecasting power of the model. In the first alternative specification we exclude the G-7 advanced countries from the sample, and in the second specification we exclude the highly indebted poor countries (HIPC).[17] These two country groupings represent opposite ends of the spectrum for the level of economic advancement. Hence, the objective is to assess whether the variables identified as important for strong growth for the whole sample are different from those when a key group of countries is excluded from the sample. Second, and more importantly, under each alternative specification we also assess to what extent the classification tree can correctly predict the growth outcomes for the excluded countries. In the final specification, we ask a slightly

[17] The definition of advanced economies and HIPC countries draw from IMF's *World Economic Outlook* regional groupings.

different question: we exclude from the list of explanatory variables those without any policy content (leaving 36 variables to consider) and ask "what are the do's and don'ts of growth?"—in other words, we first identify the top ten policies that are associated with strong growth, and then the top ten policies that are associated with poor growth (where by the dependent variable represents the bottom 25^{th} percentile of the growth performance in the sample). While we do not expect the exact same indicators that underlie strong growth performance to also underlie poor growth outcomes, the objective is to check if a particular variable is key to any kind of growth experience, good or bad.

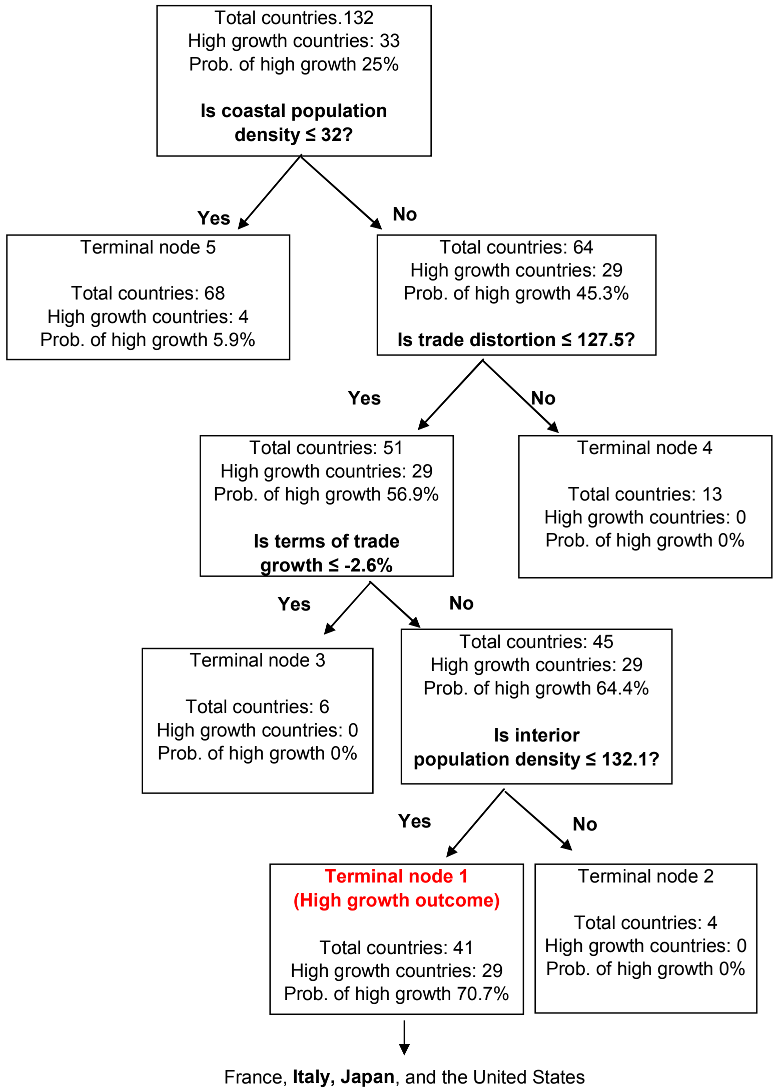
(i) Is the model appropriate for advanced economies?

In this specification, the original sample is modified to exclude the group of seven advanced economies—Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States. Among the excluded countries, Italy and Japan had strong growth performance during the sample period while the other five countries did not. Thus, the modified sample has 132 countries, with 33 strong growth performers implying that the unconditional probability of a high growth outcome is exactly at 25 percent.

Column (iii) of Table 3 shows the variable ranking with the modified sample, while Figure 3 shows the classification tree. As shown in Table 3, the variable ranking is very similar to that under the baseline—16 out of the 19 variables identified as important under this alternative specification were also significant in the baseline specification, with the first five indicators ranked identically. The ensuing tree from the alternative specification is also similar to the baseline case (Figure 3)—coastal population density is the primary splitter at the parent node, followed by trade distortion and terms of trade growth. Only the last split is now based on interior population density rather than initial level of GDP. The sequence of conditions that increases the conditional probability of a strong growth outcome from 25 percent (in the parent node) to 71 percent (in the strong growth node) is: a coastal population density higher than 32 people/km^2, trade distortion index of less than 127.5, terms of trade growth higher than -2.6 percent, and an interior population density less than 132.1 people/km^2. The terminal nodes representing missed growth outcomes are characterized by relatively low coastal population density, or high trade distortion, or large decline in terms of trade or high interior population density. Thus, as in SDM, high interior population density is negatively related to growth.

Using the results in Figure 3 we predict the growth outcome in the G7 countries. The model correctly identifies both the strong growth outcome cases (implying a type I error of zero), but incorrectly classifies 2 of the remaining 5 as strong growth performers (France and United States). Among these, France, with an average growth of 2.6 percent was very close to the threshold used to distinguish strong growth performance (2.7 percent).

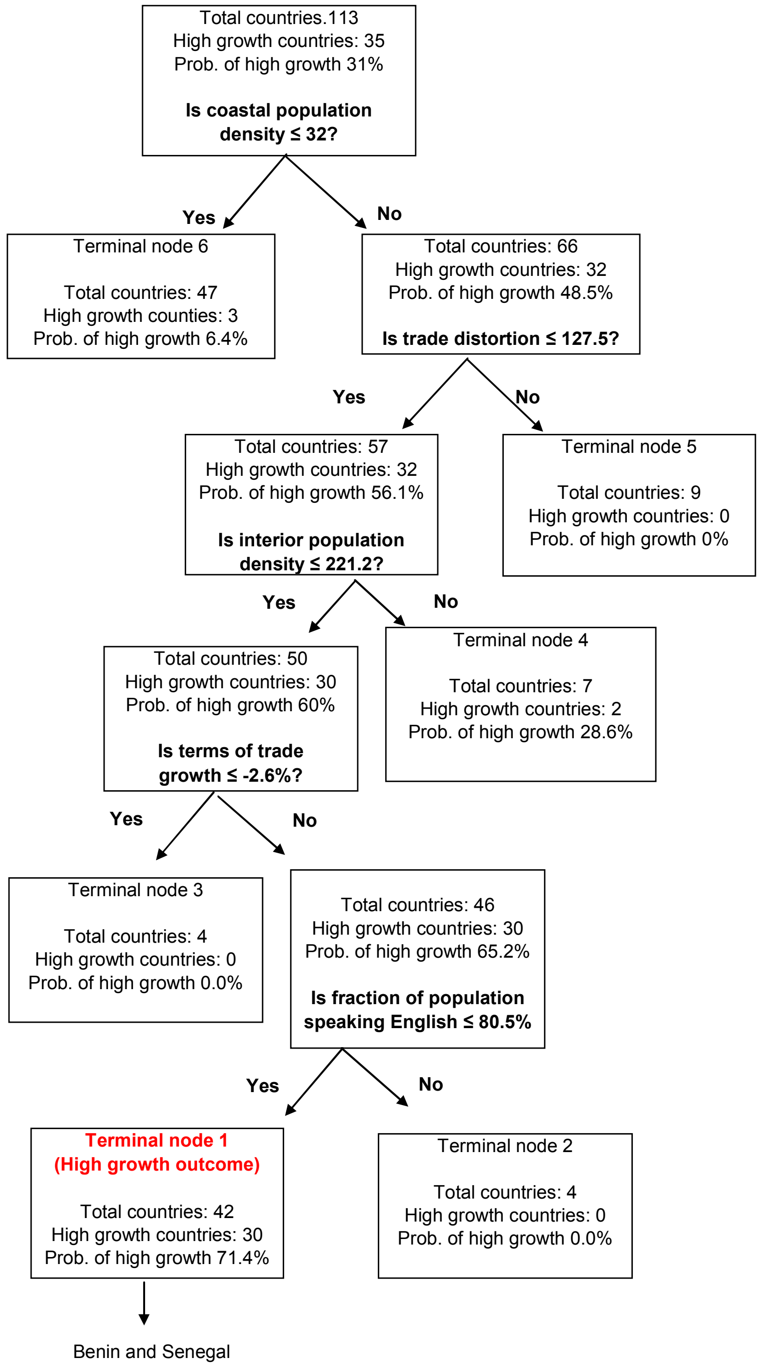
Figure 3. Out of Sample Forecast (I)—Advanced Economies



(ii) Is the model appropriate for poor countries?

In the second alternative specification, we exclude 26 HIPC countries from the sample. None of these countries had a strong growth outcome, so there is no scope for a type I error. Table 4, column (iv) shows the variable ranking under this specification, while Figure 4 describes the tree. Twenty variables are identified as important for growth, 13 of which are common to the baseline specification, with population density and coastal population density, trade distortions, and air distance to big cities continuing to rank among the top five indicators as in the baseline specification. Figure 5 is slightly different from the tree under the baseline, although the top three splitters are identical to the baseline tree. Finally, the out-of-sample forecasts for the HIPC countries, only two of the 26 countries were incorrectly predicted (as having strong growth when they did not).

Figure 4. Out of Sample Forecast (II)—Highly Indebted Poor Countries



The above tests confirm that the BCT based on the whole model applies well to low income countries as well as advanced economies, and a wide spectrum indicators—even though some of them do not appear in the tree—help deliver a strong growth outcome. The variables that are significant under all three specifications include indicators of economic policies (trade distortions, and share of public investment in GDP), and external factors (terms of trade growth), human capital development (life expectancy), geographical and structural characteristics (total and coastal population density, fraction of population in the tropics, distance between big cities, and population under 15), and initial level of income.

(iii) What are the "do's" and "don'ts" of growth?

Table 6 is a synopsis of policies associated with strong and poor growth. The first column identifies the top ten policies associated with strong growth. These include many of the key policy indicators that were identified under the baseline model as well as others. Most importantly, policies promoting external openness continue to be important (although the percentage of years in which a country is open during the sample period outranks the trade distortion index, which is ranked 15 and not shown in Table 6), as well as human capital indicators such as life expectancy, population characteristics, fertility rate and school enrollment, and the initial level of income. Besides these, government indicators such as government share in consumption and public education investment, and public defense spending are also important for strong growth.

Table 6. The "Do's" and "Don'ts" of Growth

	Key Policies Related to Strong Growth	Key Policies Related to Poor Growth
1	**Share of years open**	Malaria prevalence rate
2	**Life expectancy**	**Higher education enrollment rate**
3	**Population growth**	**Life expectancy**
4	Share of public investment in education	**Primary school enrollment rate**
5	Share of government in consumption	**Population under 15**
6	Fertility rate	**Initial per capita income**
7	**Initial per capita income**	**Share of years open**
8	**Population under 15**	Population over 65
9	Share of government defense spending in GDP	**Population growth**
10	**Higher education enrollment rate**	Terms of trade growth

The second column shows that among the top 10 policy variables key for distinguishing poor growth, 7 are similar to those important for strong growth: these are education indicators, life expectancy, time spent in openness, population characteristics, and initial level of income. In general, the relationship between poor growth and these variables is opposite in sign from the relationship between them and strong growth. Finally, while the prevalence rate for malaria is key to explaining poor growth, it is not among the top ten indicators for strong growth outcomes, i.e., while a high prevalence of malaria is associated with poor growth, the absence of this condition does not necessarily promote strong growth.[18]

[18] The classification trees using the 36 variables are not presented here, but are available upon request.

VI. CONCLUDING REMARKS

The empirical growth literature is extremely rich with studies attempting to uncover the foundations of growth, but there is still a lot of debate among the various studies. Our paper contributes to the debate on "what delivers strong growth" using the Binary Classification Tree methodology. The advantage of using the BCT approach is that it does not require any assumptions about the underlying functional form of the model and is particularly useful when theoretical underpinnings of the outcome of interest are not yet settled, as is the case in the growth literature.

Our empirical approach contributes to the literature in a number of ways. First, the BCT approach recognizes that there may not be a straightforward linear relationship between growth and its causes in that some variables may improve the chances of growth only after reaching a specific threshold, which the model identifies. Hence, a possible reason behind the observed lack of consensus between existing studies of growth could be because most of them have so far not explicitly considered this non-linear dimension between growth and the usual suspects. Second, our model recognizes the importance of conditional thresholds, in that a strong growth outcome may be delivered when a sequence of conditions or rules are satisfied. Third, as in conventional regression analysis the model also identifies which variables are robustly related to a strong growth outcome.

Our results suggest that a strong growth outcome is underpinned by a wide spectrum of economic-structural-geographical characteristics. The ten variables that are identified as having a robust relationship with growth can be consolidated into the following broad categories: (i) indicators of external sector performance (trade distortions, and terms of trade growth); (ii) government policies (share of public investment in GDP), (iii) human capital (primary school enrollment), (iv) geographical and structural characteristics (total and coastal population density, fraction of population in the tropics, distance between big cities, and population under 15), and (v) the initial level of income. These variables have an important bearing on growth for the whole sample, as well as for more restricted specifications that exclude the very advanced economies and the very poor countries from the sample. Overall, our results, in terms of identifying what is key for growth, is the closest to Sala-i-Martin and others among the empirical cross-sectional growth studies in the literature.

We identify a set of rules—proxying for external sector performance and policies, and supporting conditional convergence—that increase the chances of a strong growth by almost three times: when the coastal population density is more than 32 people/km^2, a trade distortion index less than 127.5 (where 100 implies no distortion), terms of trade growth is stronger than -2.6 percent, and the initial real per capita income is less than US$5,432. These conditions make the case that a strong growth outcome is delivered by a number of favorable outcomes occurring at the same time rather than a single factor, which is the most appealing contribution of the model.

A reasonable conclusion to draw from the foregoing is that a certain amount of geographical luck and a favorable external environment, combined with strong economic policies that promote external competitiveness and the development of human resources are really good for growth. We do not find compelling evidence that only a particular set of variables matter for growth, e.g., "institutions rule," or "it's geography, stupid," but that a certain combination of factors are important.

Finally, it is important to stress that we do not attempt here to develop a new theory of growth.[19] The evidence gathered here compels us to the conclusion that (strong) growth is a complex process that requires certain initial conditions, good socio-economic policies, and a certain dose of good luck.

[19] See for instance, Galor (2005) for a more ambitious effort at creating a unified growth theory aimed at explaining growth throughout the entire history of the human race.

APPENDIX 1. DESCRIPTION OF THE DATABASE

	Variable	Description
	Dependent variable: Average growth rate of GDP per capita 1960-1996	Growth of GDP per capita at PPP between 1960-96
	Independent variables	
1	Average inflation 1960-1990	Average inflation rate between 1960 and 1990. 1/
2	Square of inflation	Square of average inflation rate
3	Openness measure, 1965-74	Ratio of exports plus imports to GDP, averaged over 1965-74
4	Trade distortion	A trade distortion index
5	Outward orientation	A measure of outward orientation.
6	Years open 1950-1994	Number of years economy has been open between 1950 and 1994
7	Terms of trade growth in 1960s	Growth of terms of trade in the 1960's.
8	Terms of trade ranking	Ranking of countries in terms of their terms of trade
9	Fertility in 1960's	Fertility rate in 1960's.
10	Primary schooling in 1960	Enrollment rate in primary education In 1960.
11	Higher education in 1960	Enrollment rates in higher education.
12	Life expectancy in 1960	Life expectancy in 1960.
13	Malaria prevalence in 1960's	Index of malaria prevalence in 1966.
14	Population growth rate	Average growth rate of population between 1960 and 1990.
15	Population in 1960	Population in 1960.
16	Fraction population over 65	Fraction of population older than 65 years in 1960.
17	Fraction population less than 15	Fraction of population younger than 15 years in 1960.
18	Defense spending share	Average share public expenditures on defense as fraction of GDP between 1960 and 1965.
19	Public education spending share in 1960s	Average share public expenditures on education as fraction of GDP between 1960 and 1965.
20	Public investment share	Average share of expenditures on public investment as fraction of GDP between 1960 and 1965.
21	Nominal government GDP share 1960s	Average share of nominal government spending to GDP, 1960-64
22	Real government share of GDP in 1960s	Average share real government spending to GDP between 1960-1964.
23	Government consumption share (1960s)	Share of expenditures on government consumption to GDP in 1961.
24	GDP in 1960 (log)	Logarithm of GDP per capita in 1960.
25	Size of economy, 1960	Log of aggregate GDP in 1960.
26	Investment price	Average investment price level between 1960 and 1964 on PPP basis
27	Share of GDP in mining	Fraction of GDP in mining.
28	Oil-producing country dummy	Dummy for oil-producing country.
29	Share of primary exports 1970	Fraction of primary exports in total exports in 1970.
30	Civil liberties	Index of civil liberties index in 1972.
31	Political rights	Political rights index.
32	Revolutions and coups	Number of revolutions and military coups.
33	Fraction spent in war	Fraction of time spent in war between 1960 and 1990.
34	War participation 1960-1990	Indicator for countries that participated in external war during 1960-90
35	Capitalism	Capitalism index
36	Socialist dummy	Dummy for countries under Socialist rule for considerable time during 1950 to 1995.
37	Ethnolinguistic fractionalization	Average of five different indices of ethnolinguistic fractionalization which is the probability of two random people in a country not speaking the same language.
38	English-speaking population	Fraction of population speaking English. From Hall and Jones (1999).
39	Fraction speaking foreign language	Fraction of population speaking foreign language.
40	Colony dummy	Dummy for former colony.
41	British colony	Dummy for former British colony after 1776.
42	Spanish colony	Dummy variable for former Spanish colonies.
43	Timing of independence	Timing of national independence measure: 0 if before 1914; 1 if between 1914 and 1945; 2 if between 1946 and 1989; and 3 if after 1989.
44	Fraction Buddhist	Fraction of population Buddhist in 1960.
45	Fraction Catholic	Fraction of population Catholic in 1960.
46	Fraction Orthodox	Fraction of population Orthodox in 1960
47	Fraction Protestant	Fraction of population Protestant In 1960.
48	Fraction Confucian	Fraction of population Confucian.
49	Fraction Hindu	Fraction of the population Hindu in 1960.
50	Fraction Muslim	Fraction of population Muslim in 1960.
51	Religious intensity	Religion measure.
52	Absolute latitude	Absolute latitude.
53	East Asian dummy	Dummy for East Asian countries
54	European dummy	Dummy for European economies.
55	African dummy	Dummy for Sub-Saharan African countries.
56	Latin American dummy	Dummy for Latin American countries.
57	Landlocked country dummy	Dummy for landlocked countries.
58	Tropical climate zone	Fraction tropical climate zone.
59	Fraction of tropical area	Proportion of country's land area within geographical tropics.
60	Population density 1960	Population per area in 1960.
61	Population density coastal in 1960's	Coastal (within 100 km of coastline) population per coastal area, 1965
62	Fraction population in tropics	Proportion of country's population living in geographical tropics.
63	Interior density	Interior (more than 100 km from coastline) population per interior area in 1965.
64	Air distance to big cities	Log of minimal distance (in km) from New York, Rotterdam, or Tokyo.
65	Hydrocarbon deposits in 1993	Log of hydrocarbon deposits in 1993.
66	Fraction of land area near navigable water	Proportion of country's land area within 100 km of ocean
67	Land area	Area in square km.

Source: Sala-i-Martin and others (2004).

1/ This data is replaced by average inflation during 1960-69 from the *World Economic Outlook* , IMF.

REFERENCES

Acemoglu, D., S. Johnson, and J.A. Robinson, 2001, "The Colonial Origins of Comparative Development: An Empirical Investigation," *The American Economic Review*, Vol. 91, No.4, pp. 1369-1401.

Barro, R.J., 1997, Determinants of Economic Growth: A Cross-Country Empirical Study, (Cambridge Massachusetts: MIT Press).

Barro, R.J., 1991, "Economic Growth in a Cross Section of Countries," *Quarterly Journal of Economics,* Vol. 106, No. 2, pp. 407-43.

Barro, R.J., and X. Sala-i-Martin, 2004, *Economic Growth*, (New York: McGraw Hill).

Bleaney, M., and A. Nishiyama, 2002, "Explaining Growth: A Contest Between Models," *Journal of Economic Growth*, Vol. 7, pp. 43-56.

Brock, W.A., and S.N. Durlauf, 2000, "Growth Economics and Reality," NBER Working Paper No. 8041 (Cambridge, Massachusetts: National Bureau of Economic Research).

Dollar, D., 1992, "Outward-Oriented Developing Economies Really Do Grow More Rapidly: Evidence from 95 LDCs, 1976-1985," *Economic Development and Cultural Change*, Vol. 40, No. 3 (April), pp. 523-44.

Duttagupta, R., and P. Cashin, 2008, "Anatomy of Banking Crises: A Binary Classification Tree Approach," IMF Working Paper 08/93, (Washington DC: International Monetary Fund).

Easterly, W. and R. Levine, 1997, "Africa's Growth Tragedy: Policies and Ethnic Divisions," Quarterly Journal of Economics, Vol. 112, No.4, pp. 1203-50.

Galor, O., 2005, "From Stagnation to Growth: Unified Growth Theory," in Aghion and Durlauf (Ed.), *Handbook of Economic Growth,* Elsevier.

Ghosh, A., and S. Phillips, 1998, "Warning: Inflation May Be Harmful to Your Growth," *IMF Staff Papers,* Vol. 45. No.4.

Ghosh, A., and H.C. Wolf, 1998, "Thresholds and Context Dependence in Growth," NBER Working Paper No. 6480.

Leamer, E.E., 1983, "Let's Take the Con Out of Econometrics," *The American Economic Review*, Vol. 73, No.1, pp. 31-43.

Leamer, E.E., 1985, "Sensitivity Analysis Would Help," *The American Economic Review,* Vol. 75, No.2, pp. 308-13.

Levine, R., and D. Renelt, 1992, "A Sensitivity Analysis of Cross-Country Growth Regressions," *The American Economic Review*, Vol. 82, No. , pp. 942-63.

Hoover, K.D., and S.J. Perez, 2004, "Truth and Robustness in Cross-Country Growth Regressions," *Oxford Bulletin of Economics and Statistics,* Vol. 66, No. 5, pp. 765-798.

Hendry, D.F., and H.M. Krolzig, 2004, "We Ran One Regression," mimeo, Oxford University.

Reade, J., 2006, "Econometricians Behaving Badly," mimeo.

Rodrik, D., A. Subramanian, and F. Trebbi, 2004, "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development," *Journal of Economic Growth,* Vol. 9, pp. 131-65.

Romer, P.M., 1986, "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, Vol. 94, pp. 1002-37.

Sachs, J.D., 2001, "Tropical Underdevelopment," NBER Working Paper No. 8119, (Cambridge, Massachusetts: National Bureau of Economic Research).

Sachs, J.D., and A.M. Warner, 1995, "Economic Reform and the Process of Integration," *Brookings Papers on Economic Activity*, pp. 1-95

Sachs, J.D., and A.M. Warner, 1995, "Sources of Slow Growth in African Economies," *Journal of African Economies,* Vol. 6, pp. 335-76.

Sala-i-Martin, X., 1997, "I Just Run Two Million Regressions," *The American Economic Review,* Vol. 87, No. 2, pp. 178-83.

Sala-i-Martin, X., G. Doppelhofer, and R.I. Miller, 2004, "Determinants of Long-Term Growth: A Bayesian Averaging of Classical (BACE) Approach," *The American Economic Review,* Vol. 94, No. 4, pp. 813-35.

Solow, R.M., 1956, "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics,* Vol. 70, pp. 65-94.

Swan, T.W., 1956, "Economic Growth and Capital Accumulation," *Economic Record,* Vol. 32, pp. 334-61.

Wolf, H., 2004, "Accounting for Consumption Volatility Differences," *IMF Staff Papers,* Vol. 51 (Special Issue), pp 109-125.