

NBER WORKING PAPERS SERIES

LOCAL VERSUS GLOBAL CONVERGENCE ACROSS NATIONAL ECONOMIES

Steven N. Durlauf

Paul A. Johnson

Working Paper No. 3996

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 1992

Much of this paper was written while the first author was visiting the London School of Economics, whose hospitality he gratefully acknowledges. This research was supported by the Center for Economic Policy Research. We thank Robert Barro and Paul Romer for several helpful discussions and seminar participants at Cambridge, Federal Reserve Board of Governors, London School of Economics, NBER Summer Institute, NBER Fall 1991 Economic Growth Conference, Oregon and Oxford for valuable comments. We especially thank Doug Steigerwald for initially suggesting the use of regression tree methods. This paper is part of NBER's research programs in Economic Fluctuations and Growth. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #3996
February 1992

LOCAL VERSUS GLOBAL CONVERGENCE ACROSS NATIONAL ECONOMIES

ABSTRACT

This paper reexamines the ability of the Solow-type growth models to explain the pattern of cross-country growth rates. Recent authors, most notably Mankiw, Romer and Weil [1990], have argued that differences in national growth rates are compatible with the view that each country has access to a common, neoclassical aggregate production function. Such models imply that, conditional on population growth and savings rates, disparate economies are converging over time to the same level of per capita output. We argue that cross-country growth is better explained by a model of local versus global convergence. Countries converge locally in the sense that economies with similar initial conditions tend to converge to one another. However, we find little evidence of convergence across economies with substantially different initial conditions as measured by per capita output or literacy rates. Further, the impact of capital formation on aggregate output increases with the level of economic development. These results are consistent with models of multiple equilibria in long run behavior. Our results suggest that the Solow growth model should be supplemented with a theory of aggregate production function differences in order to fully explain international growth patterns.

Steven N. Durlauf
Department of Economics
Stanford University
Stanford, CA 94305-6072
and NBER

Paul A. Johnson
Department of Economics
University of Oregon
Eugene, OR 97403

Introduction

Starting with Baumol [1986], a number of authors have explored the behavior of output growth across different aggregate economies. These authors have generally been interested in understanding whether economies exhibit convergence—defined as a tendency for per capita output to equalize over time. This question has important implications for the utility of various theoretical growth models. In particular, the neoclassical growth model developed by Solow predicts that different economies will converge in the sense that per capita output differences due to initial conditions will asymptotically disappear as economies are assumed to have access to identical concave production technologies. The new growth theory pioneered by Romer [1986], Lucas [1988], and Grossman and Helpman [1991], on the other hand, shows how various types of production nonconvexities can interact with market imperfections to produce multiple long run output equilibria for a given microeconomic specification, which means that per capita output differences can be persistent.

Much of the empirical work on convergence has been concerned with determining whether poor economies grow faster than wealthier ones, which is equivalent to identifying a negative cross-section correlation between a country's initial per capita output and subsequent growth rate for a fixed period afterwards. Formally, if $(Y/L)_{i,T}$ equals the per capita output of country i at T , convergence across a set of N countries is said to occur over a fixed epoch τ if the coefficient β in the cross-section regression

$$\ln(Y/L)_{i,T+\tau} - \ln(Y/L)_{i,T} = \zeta + \beta \ln(Y/L)_{i,T} + \Pi \tilde{X}_i + \epsilon_i, \quad i = 1 \dots N \quad (1)$$

is negative. Here, \tilde{X}_i denotes a set of control variables with associated coefficients Π , usually meant to control for microeconomic heterogeneity. A negative correlation is generally necessary for income differences to narrow, although as shown in Bernard and Durlauf [1991], the condition is far from sufficient.

Overall, the evidence on convergence is somewhat mixed, with results depending on both the sample of countries studied as well as the choice of control variables. One robust result is that in the absence of any control variables, a negative β can be found for OECD economies, as demonstrated by Baumol [1986] among others, whereas the value of β is typically zero or even slightly positive when cross-section regressions are run for a large country data set such as the one developed by Summers and Heston [1988], as shown by many authors. Second, it is clear that there exist many plausible choices of X_i ; control variables which can extend the convergence results to a wide cross-section of economies. For example, Barro [1991] shows how controlling for education, investment rates and political stability, among other variables, can allow a negative β to emerge for the countries in the Summers-Heston data set.

One of the most careful and provocative of the pro-convergence studies is due to Mankiw, Romer and Weil [1990] who study a cross-section regression where the control variables are not *ad hoc* additions to the equation but rather are directly suggested by the law of motion for per capita output produced by a human capital-augmented version of the Solow model. These variables control for the rates of savings of human and physical capital and the rates of population growth, technical change and depreciation. These authors not only find that there is strong evidence of convergence for a broad country sample, but also conclude that nearly 50% of the cross-country variation in growth can be attributed to the Solow model.

One difficulty with the body of cross-section studies is that they often do not make clear the nature of the null and alternative models associated with a particular statistical test. For example, does a negative β in a regression restricted to a group of advanced industrialized economies such as the OECD represent evidence supportive of the Solow model as opposed to the Romer-Lucas class of models? The answer is no. Many new growth models, such as Azariadis and Drazen [1990], imply the existence of several locally stable equilibria in long run per capita output. In their framework, the OECD could represent a group of economies which are converging to a (relatively) high

production equilibrium. Such multiple equilibrium models predict that one will observe convergence once one has isolated economies which are associated with the same equilibrium. In this sense, restrictions of the cross-country sample under consideration can lead to spurious inferences with respect to competing growth theories.

Further, as argued in Bernard and Durlauf [1991], it is not clear that the estimation of (1) will fail to produce a negative β for samples where convergence does not hold. To see this, suppose that there exist M different long run equilibria for economies. In addition, suppose that the "correct" model of the evolution of these economies is

$$\ln(Y/L)_{i,T+\tau} - \ln(Y/L)_{i,T} = \zeta_i + \beta \ln(Y/L)_{i,T} + \Pi X_i + \epsilon_i, \quad i = 1, \dots, N \quad (2)$$

where ζ_i is determined by the country's long run equilibrium. In this case, the law of motion for each economy is the same except for the constant term. A version of the capital complementarities model described in Romer [1986] can be shown to obey an equation of this type. The use of ζ rather than of ζ_i in the regression will, of course, bias the coefficients. However, this misspecified model will reject convergence only if β becomes nonnegative, which will depend in a complicated way on the covariance structure of ζ_i with the remaining right hand side variables. Bernard and Durlauf [1991] in fact show that a negative β is compatible with a nontrivial class of multiple equilibrium/endogenous growth models.

Similarly, one can also see how the addition of control variables to the basic equation (1) can spuriously induce a negative β for data generated by a multiple equilibrium model. Proxying for the missing ζ_i , control variables can have the effect of permitting the regression to separate countries into subgroups. For example, suppose that there are two equilibria for aggregate economies, which can be distinguished by different mean levels of growth. If some control variable x_i is added to equation (1) which positively correlates with the growth rates across economies (such as per capita years of schooling), the variable can act to segregate the data into different regimes and

again lead to a negative estimate of β .

The purpose of this paper is to reexamine the cross-section behavior of growth rates in the Summers-Heston data set to see in what sense the data are supportive of convergence. We do this by distinguishing local convergence from global convergence. By local convergence, we refer to the case where there exist groups of countries in the Heston-Summers data set such that convergence occurs within the groups. Global convergence holds when all economies converge to one another.

Our analysis takes the regressions studied by Mankiw, Romer and Weil (which we shall designate as M-R-W) as a starting point and asks whether the same data are compatible with, and indeed better (in a standard goodness-of-fit sense) modeled as realizations from a model with multiple equilibria. We choose the M-R-W paper because we regard it as the most compelling evidence in favor of the Solow model in the literature. We do not dispute their conclusion that the Solow model has substantial explanatory power for cross-country variation in growth rates. What we do is show that extending the Solow model to allow for additional factors, in particular aggregate production function differences, is important. In fact, by segregating countries into locally converging groups which allow for different aggregate production functions, we find that the Solow model has substantially more explanatory power than has been previously suggested.

Our empirical conclusions are twofold. First, we provide evidence that there exist groups of countries exhibiting local rather than global convergence. Mechanically splitting the Summers-Heston data by either initial income or initial literacy generates a statistically significant improvement in the ability of a Solow-type model to explain growth rates within each group. We reject the null hypothesis that the groups of economies are converging to one another. Second, by applying both maximum likelihood and regression tree procedures to the data in order to endogenously identify convergence groups, we find that different countries in the data set obey very different production functions. Our analysis finds that the share of labor in total income tends to decline with

the level of development as measured by initial income levels or literacy rates, implying that, other things equal, more developed countries will have higher output/labor ratios than implied by their capital/labor ratios alone. These results verify the idea of Baumol [1986] that there may be several "convergence clubs" in the world economy.

Methodologically, these results illustrate the sorts of specification and estimation exercises which need to be applied in growth studies in order to identify the uniqueness versus multiplicity of long run equilibria. Empirical work which seeks to assess those theories necessarily must attempt to identify different data regimes in order to ensure testing power. Our specification tests provide a way of overcoming the low power of standard cross-section tests as documented in Bernard and Durlauf [1991] and our use of maximum likelihood and regression tree methods illustrates a straightforward way of identifying the different regimes.

Section 1 reviews the link between standard growth models and the cross-section regressions we study. Section 2 describes the data we analyze. Section 3 performs some specification tests on cross-country regressions estimated on the Summers-Heston data. Using initial income and literacy rates to segregate the countries, we reject the null hypothesis that the data come from a single regime. We also check the robustness of our results for some different formulations of the single regime model. Section 4 uses maximum likelihood and regression tree techniques to endogenously identify groups of locally converging economies. The analysis allows for the interaction of different control variables in determining data regimes. Section 5 provides summary and conclusions. Data and Technical Appendices follow.

1. Convergence and cross-section behavior

In this section, we illustrate two senses in which cross-section regressions of the

form (1) are appropriate tests of convergence. Following the exposition of M-R-W, we define the following variables which shall also be used in the empirical section. The subscripts i and t index countries and time periods respectively.

A_t = level of technology at t , assumed constant across i .

$L_{i,t}$ = labor input.

$Y_{i,t}$ = aggregate output.

$K_{i,t}$ = physical capital input.

$H_{i,t}$ = human capital input.

n_i = population growth rate, assumed constant across t .

g = rate of technical change, assumed constant across i and t .

δ = depreciation rate of physical and human capital, assumed constant across i and t .

s_i^k = savings rate for physical capital, assumed constant across t .

s_i^h = savings rate for human capital, assumed constant across t .

Equation (1) can be justified first as the law of motion generated by the Solow growth model. To see this, we consider the case where the production function is Cobb-Douglas,

$$Y_{i,t} = \phi K_{i,t}^\alpha H_{i,t}^\gamma (A_t L_{i,t})^{1-\alpha-\gamma}. \quad (3)$$

Labor and the level of technology each grow exponentially,

$$L_{i,t} = L_{i,0} e^{n_i t} \quad (4)$$

and

$$A_t = A_0 e^{gt}. \quad (5)$$

The evolution of physical and human capital per labor input are similarly determined by the interaction of the exogenously determined physical and human capital savings rates with depreciation.

$$\frac{dK_{i,t}}{dt} = s_i^k Y_{i,t} - \delta K_{i,t} \quad (6)$$

$$\frac{dH_{i,t}}{dt} = s_i^h Y_{i,t} - \delta H_{i,t} \quad (7)$$

These laws of motion for technology and the various production inputs lead to a law of motion for output per worker, $(Y/L)_{i,t}$, over any interval T to $T+\tau$.

$$\begin{aligned} \ln(Y/L)_{i,T+\tau} - \ln(Y/L)_{i,T} = \\ g\tau + (1 - e^{-\lambda_i \tau}) \left(\Theta + \frac{\alpha}{1 - \alpha - \gamma} \ln(s_i^k) + \frac{\gamma}{1 - \alpha - \gamma} \ln(s_i^h) \right) \\ - \frac{\alpha + \gamma}{1 - \alpha - \gamma} \ln(n_i + g + \delta) - \ln(Y/L)_{i,T} \end{aligned} \quad (8)$$

Here, $\Theta = \frac{1}{1 - \alpha - \gamma} \ln(\phi) - \ln(A_0) - gT$ and $\lambda_i = (1 - \alpha - \gamma)(n_i + g + \delta)$, the country-specific convergence rate towards the steady state.

Equation (8) thus provides a way of explaining cross-country growth rates through a common technology and country-specific input growth. Observe that, if we impose the restriction $\lambda_i = \lambda$ for all i , as M-R-W do, the equation takes on the form of equation (1). In this case, the coefficients comprising Π are functions of the structural parameters α and γ , which allows one to consider both unconstrained and constrained versions of (8). However, since $\lambda_i = (1 - \alpha - \gamma)(n_i + g + \delta)$, the restriction of equal convergence rates across countries is valid only if population growth occurs at a constant

rate in each country. This is far from the case – in our sample the coefficient of variation of population growth is similar to that of other variables.¹ In the absence of the restriction $\lambda_i = \lambda$, equations (1) and (8) are not nested. One can construct an alternative equation which nests them both but we use equation (1) because of its relationship to the work of others. We refer to equation (1) as “unconstrained” and equation (8) as “constrained” for ease of reference.²

A second justification for the cross-section regression is as follows. Assume that we wish to test whether contemporaneous output differences tend to narrow, i.e. whether the conditional expectation of the difference in output between economies i and j over a fixed horizon τ is smaller than the contemporaneous disparity. For information set \mathfrak{F}_T , this means that

$$E(\ln(Y/L)_{i,T+\tau} - \ln(Y/L)_{j,T+\tau} | \mathfrak{F}_T) < \ln(Y/L)_{i,T} - \ln(Y/L)_{j,T}. \quad (9)$$

Suppose the growth of output obeys the relation

$$E(\ln(Y/L)_{i,T+\tau} - \ln(Y/L)_{i,T} | \mathfrak{F}_T) = \zeta + \beta \ln(Y/L)_{i,T} \quad (10)$$

In this case, a negative β is necessary for (9) to hold across a set of economies. Notice that if the conditional expectation of output changes depends on more than initial income, i.e. equation (10) is misspecified, then the cross-section regression generally ignores information in assessing whether convergence holds.

The behavior of the cross-section regression (1) when the sample contains

¹Using the data described below, the coefficient of variation of the average rate of population growth over the period 1960-1985 is .404. Those for the average rate of GDP per capita growth over the same period and our measures of physical and human capital accumulation are .465, .448, and .643, respectively.

²Equation (8) does, of course, impose more restrictions than equation (1) on the model that nests them both.

diverging economies is, of course, extremely sensitive to the specific alternative under question. We outline the implications of two possible specifications of the aggregate production function which fail to generate convergence even under the Solow savings specification.

Much of the work on alternatives to the Solow model has argued that there exists a region of capital values over which the aggregate production function is not concave, which will lead to different long run steady states for different initial conditions.³ Following Romer [1986], this could occur if social increasing returns to scale are present in the economy due to effects such as learning-by-doing. To formalize this alternative in the context of the Solow model, let the production function obey

$$Y_{i,t} = \phi \xi(K_{i,t}) K_{i,t}^\alpha H_{i,t}^\gamma (A_t L_{i,t})^{1-\alpha-\gamma} \quad (11)$$

where $\xi(\cdot)$ is a continuous, nondecreasing function such that

$$\xi(K_{i,t}) = \begin{cases} \underline{\xi} & \text{if } K_{i,t} < \bar{K} \\ \bar{\xi} & \text{if } \bar{K} < K_{i,t} \end{cases} \quad (12)$$

with $\underline{\xi} < \bar{\xi}$. It is straightforward to show how a suitable choice of $\xi(\cdot)$ will induce two locally stable equilibria. From the perspective of the cross-section regressions, data generated by economies associated with a particular equilibrium will obey equation (8). A regression mixing economies from the two equilibria will be misspecified in the sense that the parameter Θ in (8) will be a function of a given economy's equilibrium. The

³Many multiple equilibrium models imply the existence of aggregate production function differences which distinguish underdeveloped and developed countries. Murphy, Shleifer and Vishny [1989] equate industrialization with the conversion of an economy to more efficient production; Durlauf [1991] shows how industrialization can occur through the build up of localized technological complementarities which expand aggregate productivity.

correctly specified cross-section law of motion takes the form

$$\begin{aligned} & \ln(Y/L)_{i,T+\tau} - \ln(Y/L)_{i,T} = \\ & g\tau + (1 - e^{-\lambda_i\tau}) \left(\Theta_i + \frac{\alpha}{1-\alpha-\gamma} \ln(s_i^k) + \frac{\gamma}{1-\alpha-\gamma} \ln(s_i^h) \right) \\ & - \frac{\alpha+\gamma}{1-\alpha-\gamma} \ln(n_i + g + \delta) - \ln(Y/L)_{i,T} \end{aligned} \quad (13)$$

where $\Theta_i = \frac{1}{1-\alpha-\gamma} \ln(\phi \underline{\xi}) - \ln(A_0) - gT$ if economy i is associated with the low production equilibrium and $\Theta_i = \frac{1}{1-\alpha-\gamma} \ln(\phi \bar{\xi}) - \ln(A_0) - gT$ if the economy is associated with the high production equilibrium.

A similar law of motion is generated by the model of human capital threshold externalities explored by Azariadis and Drazen [1990]. Azariadis and Drazen argue that there may exist human or physical capital accumulation thresholds which identify shifts in aggregate technology. For example, some minimum level of human capital per worker may be required to permit the use of more advanced technologies. One way to model this idea is to posit the existence of a human capital threshold \bar{H} such that

$$Y_{i,t} = \begin{cases} \phi K_{i,t}^\alpha H_{i,t}^\gamma (A_t L_{i,t})^{1-\alpha-\gamma} & \text{if } H_{i,t} < \bar{H} \\ \psi K_{i,t}^\eta H_{i,t}^\zeta (A_t L_{i,t})^{1-\eta-\zeta} & \text{if } H_{i,t} \geq \bar{H} \end{cases} \quad (14)$$

Again, this type of nonconvex technology will generate multiple equilibria when combined with exogenous savings rates. As before, there is no necessary implication that a cross-section growth regression using data from this model will produce a nonnegative β . However, in this case, a switching regimes analysis of cross-section data will produce different coefficients for all of the variables found in equation (8), which allows for a more general alternative to a single law of motion.

The cross-section regression (1) is correctly specified for subsets of countries when the aggregate technology obeys (13) or (14). In fact, under our versions of the Romer and Azariadis-Drazen specifications, the Solow model holds locally even though the aggregate production technology exhibits nonconvexity over some range. Against this class of alternatives, the testable implications of the Solow model are summarized in the requirement that the cross-section data are generated by a common law of motion for all countries.

2. Data

All cross-country growth rates we employ are based upon the Summers-Heston [1988] international output estimates. With the exception of the data on literacy rates, which are taken from the World Bank's *World Development Report* (various issues), all of the data that we use are from M-R-W.⁴ The variables are defined as follows:

$(Y/L)_{i,t}$ = real GDP per member of the population aged 15-64, country i at t .

$(I/Y)_i$ = fraction of real GDP devoted to investment (including government investment), country i , annual average for 1960-1985.

n_i = growth rate of the working-age population, country i , annual average 1960-1985.

$SCHOOL_i$ = fraction of the working-age population enrolled in secondary school, country i , annual average 1960-1985.

⁴The primary sources are the data set constructed by Summers and Heston [1988] and the World Bank's *World Tables* and *World Development Report*. M-R-W discuss the construction and some of the limitations of the data. A Data Appendix at the end of the paper lists the 98 countries in our data set as well as some selected characteristics.

$LR_{i,1960}$ = adult literacy rate, fraction of the population aged 15 and over that is able to read and write, country i , in 1960.⁵

We follow M-R-W in assuming that $g = 0.02$ (implying that $gr = 0.5$, a value that we impose in estimation) and $\delta = 0.03$, figures that are approximately true for the United States. We also follow these authors in using $(I/Y)_i$ to represent s_i^k and $SCHOOL_i$ to represent s_i^h .

3. Specification tests for multiple regimes

In this section we attempt to identify separate regimes in the data through the use of specification tests which take a single regime model as the null hypothesis. We do this by mechanically splitting the data into subgroups based upon different control variables and examining whether model parameters are equal across groups. We consider two estimating equations. First, we fit

$$\begin{aligned} \ln(Y/L)_{i,1985} - \ln(Y/L)_{i,1960} = & \zeta + \beta \ln(Y/L)_{i,1960} \\ & + \pi_1 \ln(I/Y)_i + \pi_2 \ln(n_i + g + \delta) + \pi_3 \ln(SCHOOL)_i + \epsilon_i \end{aligned} \quad (15)$$

⁵For some countries the 1960 literacy rate is unavailable so the 1975 rate is used instead. As most of these have literacy rates of 90% or greater this has little effect on our results. In addition, for many countries, the "1960" literacy rate is actually calculated for some (unknown) year between 1958 and 1962. It seems unlikely that literacy changes by very much in a two year period so the magnitude of the resultant measurement error is probably small. Also, since we use the literacy rates to classify countries rather than as a regressor, the importance of small measurement errors is minimal. Two of the countries studied by M-R-W, Botswana and Mauritius, are omitted due to lack of data on literacy.

Table 1
Specification Tests for Different Regimes

Subsamples defined by	Unconstrained Regressions	Constrained Regressions
2-Way Split based on		
$(Y/L)_{i,1960}$	0.009	0.218
$LR_{i,1960}$	0.011	0.112
3-Way Split based on		
$(Y/L)_{i,1960}$	0.029	0.011
$LR_{i,1960}$	0.404	0.000
4-Way Split based on both		
$LR_{i,1960}$ and $(Y/L)_{i,1960}$	0.000	0.000

This table shows the marginal significance levels for the Wald tests of null hypothesis that the parameters of the indicated models are constant across the indicated subsamples. Splits are described in the text.

by least squares over each subgroup. We refer to this equation as the unconstrained version of the Solow model since we use equation (8) to determine the right-hand side variables without restricting the regression coefficients as implied by (8). We separately estimate a constrained version of the model by imposing these restrictions.

We consider two different control variables to group countries with similar characteristics. The first variable we employ is per capita output at the beginning of the sample period, $(Y/L)_{i,1960}$. Most models of multiple long run equilibria predict that if economies are concentrated around several equilibria, then their initial per capita output levels will fall into nonoverlapping categories. Second, we examine sample splits based upon the adult literacy rate of each country in 1960. The use of literacy as a segregating variable makes sense if one thinks of the potential regimes in the data as stemming from a broad notion of social and economic development.⁶

Table 1 reports the results for several different data splits. Each entry represents the significance level of a Wald test of the null hypothesis that all parameters are equal across the subsamples under analysis.⁷ The first panel of the Table divides countries into two equal sized groups by segregating high and low initial income and initial literacy countries into separate categories. Each subgroup thus consists of 48 countries. The second panel divides countries into three equal groups of 32 according these variables. The third panel allows interactions between the variables. In this case, we divide countries according to whether they lie in the high or low half of the sample according to our two controls. This segregation results in four categories: high income/high literacy (42 countries) high income/low literacy (6 countries), low income/high literacy (6 countries) and low income/low literacy (42 countries).⁸

⁶See Rauch [1989] for corroborating evidence of literacy-based regime differences.

⁷Following Barro [1991] and others, we use heteroskedasticity-corrected test statistics and standard error estimates, (see White [1980]), in order to allow for different error variances for observations from different countries. White's [1980] heteroskedasticity test reveals some evidence against a homoskedastic null. Assuming homoskedasticity in the computation of the Wald statistics increases the number of rejections of the single regime model.

Table 2
 Cross Section Regressions
 Income and Literacy-Based Sample Breaks
 Dependent Variable: $\ln\left(\frac{Y}{L}\right)_{i,1985} - \ln\left(\frac{Y}{L}\right)_{i,1960}$

M-R-W	$\left(\frac{Y}{L}\right)_{i,1960} < 1950$ and $LR_{i,1960} < 54\%$		$1950 \leq \left(\frac{Y}{L}\right)_{i,1960}$ and $54\% \leq LR_{i,1960}$	
	98	42	42	
Unconstrained Regressions				
constant	3.04 [†] (0.831)	1.40 (1.85)	0.450 (0.723)	
$\ln\left(\frac{Y}{L}\right)_{i,1960}$	-0.289 [†] (0.062)	-0.444 [†] (0.157)	-0.434 [†] (0.085)	
$\ln\left(\frac{I}{P}\right)_i$	0.524 [†] (0.087)	0.310 [†] (0.114)	0.689 [†] (0.170)	
$\ln(n + g + \delta)_i$	-0.505 (0.288)	-0.379 (0.468)	-0.545 (0.283)	
$\ln(SCHOOL)_i$	0.233 [†] (0.060)	0.209 [†] (0.094)	0.114 (0.164)	
\bar{R}^2	0.46	0.27	0.48	
σ_ϵ	0.33	0.34	0.30	
Constrained Regressions				
θ	-2.56 ^{††} (1.14)	2.29 (1.17)	-0.395 (1.24)	
α	0.431 [†] (0.061)	0.275 [†] (0.097)	0.509 [†] (0.098)	
γ	0.241 [†] (0.046)	0.217 [†] (0.061)	0.108 (0.094)	
\bar{R}^2	0.42	0.28	0.50	
σ_ϵ	0.34	0.34	0.29	

† denotes significance at asymptotic 5% level

†† This equation has been reestimated under the restriction $\lambda_i = (n_i + g + \delta)(1 - \alpha - \beta)$, where λ_i is the rate of convergence toward the steady state. This restriction was not imposed by M-R-W. Their estimates are constant = 2.46 (0.48); $\alpha = 0.48$ (0.07); $\gamma = 0.23$ (0.05); $\bar{R}^2 = 0.46$; and, $\sigma_\epsilon = 0.33$.

As the Table indicates, we find substantial evidence that the laws of motion for growth within each subgroup are different. For three of the four initial income splits, equality of coefficients across the groups is rejected at the 3% level. When initial literacy represents the control variable, we reject in two of the four cases at about 1%. Further, we overwhelmingly reject the equality of regime parameters for both unconstrained and constrained regressions based on the interactive four regime specification. This change in the significance level of the tests indicates the importance of allowing both variables to identify separate data regimes.

Table 2 reports the original M-R-W regression along with estimates of the regressions associated with the high initial income/high initial literacy and low initial income/low initial literacy splits described above. (The high initial income/low initial literacy and low initial income/high initial literacy splits are omitted due to lack of degrees of freedom.) Several of the subsample coefficients are substantially different from both one another and from the M-R-W regression. For the unconstrained regressions, the coefficient on initial income, $\ln(Y/L)_{i,1960}$, is approximately equal for the high literacy/high income and low literacy/low income groups at -.434 and -.444 respectively; these estimates are much larger than the -.289 estimate for the whole sample. This difference reveals a faster convergence rate for the subsamples than suggested for the single regime. Further, the $\ln(I/Y)_i$ coefficient for high income/high literacy countries is .689, which is over twice as large as the .310 estimate for the low income/low literacy countries and over 25% higher than the .524 estimate for the whole sample. Similarly, the implied physical capital share in output for the constrained regressions is far larger for the high income/literacy countries at .509 than for the low literacy/income countries at .275, and somewhat larger than the .431 share for the whole sample. Conversely, the low income/literacy countries exhibit a much larger coefficient for the human capital

⁸The initial two-way income splits are based on $(Y/L)_{i,1960} < \$1950$ and $\$1950 \leq (Y/L)_{i,1960}$; the three way splits are based on $(Y/L)_{i,1960} < \$1150$, $\$1150 \leq (Y/L)_{i,1960} \leq \2750 and $\$2750 < (Y/L)_{i,1960}$. For initial literacy, the two-way splits are based on $LR_{i,1960} < 54\%$ and $54\% \leq LR_{i,1960}$; the three way splits are based on $LR_{i,1960} < 26\%$, $26\% \leq LR_{i,1960} \leq 72\%$ and $72\% < LR_{i,1960}$.

investment measure $\ln(SCHOOL)_i$, as well as the associated human capital output share than high income/literacy countries, although both subsample estimates are below those for the whole sample. These estimates suggest that the aggregate production functions are substantially different across subsamples, which supports a multiple regime perspective.

Robustness and additional control variables

One explanation of these results is that the set of control variables dictated by the Solow model is too small to account for some important differences in growth performance so that our evidence of multiple regimes is actually due to omitted variables. In this case, inclusion of these variables among the X_i would render the specification correct and eliminate the statistical significance of the sample splits.

Barro [1991] uses a broader set of control variables than M-R-W in an attempt to model a wide variety of potential influences on growth. We therefore investigate whether our rejection of the single equilibrium model is robust to the addition of some of Barro's variables to those dictated by the strict Solow model. We focus on the role of government spending and education variables in Barro's work.⁹ The variables are:

$(G^C/Y)_i$ = ratio of government consumption to GDP, country i , annual average for 1960-1985 (or the largest available subperiod).

$PRIM60_i$ = primary-school enrollment rate, country i , 1960.

⁹We choose these particular variables because they represent natural extensions of the Solow model, as $(G^C/Y)_i$ can proxy for taxes and $PRIM60_i$ and $SEC60_i$ can act as additional proxies for the human capital savings rate. We have also verified that our rejections of the single regime model are robust with respect to including other variables employed by Barro; this work is available upon request.

Table 3
Specification Tests: Robustness Check

Additional regressor: $\ln(G^C/Y)_i$

Subsamples defined by	Unconstrained Regressions	Constrained Regressions
2-Way Split based on		
$(Y/L)_{i,1960}$	0.179	0.002
$LR_{i,1960}$	0.069	0.701
3-Way Split based on		
$(Y/L)_{i,1960}$	0.051	0.101
$LR_{i,1960}$	0.707	0.493
4-Way Split based on both		
$LR_{i,1960}$ and $(Y/L)_{i,1960}$	0.000	0.000

Additional regressors: $\ln(PRIM60)_i$ and $\ln(SEC60)_i$

Subsamples defined by	Unconstrained Regressions	Constrained Regressions
2-Way Split based on		
$(Y/L)_{i,1960}$	0.090	0.002
$LR_{i,1960}$	0.027	0.375
3-Way Split based on		
$(Y/L)_{i,1960}$	0.106	0.008
$LR_{i,1960}$	0.263	0.000
4-Way Split based on both		
$LR_{i,1960}$ and $(Y/L)_{i,1960}$	0.000	0.000

This table shows the marginal significance levels for the Wald tests of null hypothesis that the parameters of the indicated models are constant across the indicated subsamples. Splits are described in the text.

$SEC60_i$ = secondary-school enrollment rate, country i , 1960.¹⁰

In order to assess the impact of additional control variables, we projected $\ln(Y/L)_i, 1985 - \ln(Y/L)_i, 1960$ and the explanatory variables in equation (15) on various subsets of the logarithms of these variables (and a constant) and then used the respective residuals from these projections in place of the variables in equation (15) in calculating the test statistics for the splits described in Table 1.

Table 3 gives the results. As the Table indicates, adding $\ln(G^C/Y)_i$ as an additional regressor has some effect on the significance of the various splits. Only one of the four splits by income is now significant, although the three-way unconstrained regression is only marginally insignificant. Further, the significance of the literacy splits has been eliminated. The addition of $\ln(PRIM60)_i$ and $\ln(SEC60)_i$ as controls has less effect on the hypothesis tests. All of the income splits are significant at the 11% level and two are significant at the 1% level. Two of the four literacy splits are still significant at 3%. On the other hand, the strong significance of the four-way interactive splits is unaffected by any variable additions. The evidence of multiple regimes therefore seems robust to the addition of these variables, although the government spending variable reduces the significance of some data splits.

4. Properties of local convergence groups

In this section, we study the behavior of unconstrained and constrained growth regressions for different country groups. Although the exogenously imposed data splits of the previous section permit straightforward specification testing, they do not address the

¹⁰ $SEC60_i$ differs from $SCHOOL_i$ as it measures the ratio of secondary students to the population between 12-17 rather than to all working age persons and because it equals a point estimate for 1960 rather than an average over 1960-1985.

Table 4
 Cross Section Regressions
 Income-Based Sample Breaks
 Dependent Variable: $\ln\left(\frac{Y}{L}\right)_{i,1985} - \ln\left(\frac{Y}{L}\right)_{i,1960}$

	$\left(\frac{Y}{L}\right)_{i,1960} < 800$	$800 \leq \left(\frac{Y}{L}\right)_{i,1960} \leq 4850$	$4850 < \left(\frac{Y}{L}\right)_{i,1960}$
Observations	14	63	21
	Unconstrained Regressions		
constant	3.46 (2.27)	-0.86 (1.38)	-7.26 [†] (1.59)
$\ln\left(\frac{Y}{L}\right)_{i,1960}$	-0.791 [†] (0.269)	-0.172 (0.109)	0.069 (0.139)
$\ln\left(\frac{I}{Y}\right)_i$	0.314 [†] (0.109)	0.449 [†] (0.120)	0.475 [†] (0.119)
$\ln(n + g + \delta)_i$	-0.429 (0.678)	-0.322 (0.456)	-1.75 [†] (0.270)
$\ln(SCHOOL)_i$	-0.028 (0.073)	0.331 [†] (0.086)	0.341 [†] (0.141)
\bar{R}^2	0.57	0.52	0.82
σ_e	0.16	0.33	0.12
	Constrained Regressions		
θ	4.107 [†] (0.552)	-3.88 (2.04)	-11.0 (7.64)
α	0.306 [†] (0.083)	0.401 [†] (0.089)	0.333 [†] (0.100)
γ	-0.034 (0.083)	0.302 [†] (0.069)	0.455 [†] (0.103)
\bar{R}^2	0.64	0.47	0.71
σ_e	0.19	0.36	0.18

† denotes significance at asymptotic 5% level

problem of accurately identifying economies with similar laws of motion. In order to identify economies with similar characteristics, it is necessary to allow the data to endogenously determine the location of the different regimes. We perform two sets of exercises to identify blocks of locally converging economies.

Single control variable estimates

In order to allow the data to endogenously determine which countries belong in which group, we first employ an approach suggested by Quandt [1958]. This method consists of exogenously choosing the number of splits and then choosing their location according to a control variable so as to maximize the likelihood function of the data. We choose the number of splits to be three. Let z_i be the variable used to split the sample. For any two fixed numbers, $\underline{z} < \bar{z}$, the sample can be split into three subsamples according to whether $z_i < \underline{z}$, $\underline{z} \leq z_i \leq \bar{z}$, or $\bar{z} < z_i$. Let N_j , $j = 1, 2, 3$, be the number of observations in each of the respective subsamples. The maximized log quasi-likelihood is given by $\mathcal{L}(\underline{z}, \bar{z}) = -\sum_{j=1}^3 N_j \ln(\hat{\sigma}_j^2)$, where $\hat{\sigma}_j^2 =$ residual sum of squares divided by N_j based on estimating equation (15) on subsample j . We choose the regime split (\underline{z}, \bar{z}) that maximizes $\mathcal{L}(\underline{z}, \bar{z})$, subject to the feasibility of least squares estimation for each subsample.

Using initial income as the control variable, the likelihood function is maximized for regimes identified by the splits $(Y/L)_{i,1960} < \$800$, $\$800 \leq (Y/L)_{i,1960} \leq \4850 , and $\$4850 < (Y/L)_{i,1960}$.¹¹ Table 4 reports the unconstrained and constrained regressions for the estimated regimes. In terms of overall fit, we find some improvement over the single

¹¹See the Data Appendix for the classification of each country into estimated high, intermediate, and low initial literacy and income groups. We have also considered whether the endogenous splits are statistically significant by computing (through Monte Carlo methods) the distribution of the *sup* of the Wald statistics over all possible 3-way income and literacy splits when there is one regime and comparing the Wald statistics for our estimated splits to these distributions. We find that the income and literacy splits are significant at 3% and 2% respectively.

regime specification. Whereas M-R-W found that they could explain 46% of overall growth variation in the unconstrained model, we find that for the poorest economies, we explain 57%, for intermediate economies 52%, and for high income economies fully 82% of the total growth variation. Similar results hold for the constrained regressions.

In terms of the individual coefficients, clear evidence exists for regime-dependent sensitivity of growth to the different control variables. The coefficients on $\ln(I/Y)_i$ are .314, .449, and .475 for the low, intermediate, and high income groups respectively. The effect of the variable on growth is thus over 30% smaller for low income economies when compared to their intermediate and high income counterparts. Further, the coefficient estimates for $\ln(SCHOOL)_i$, -.028, .331, and .341, for low, intermediate, and high income groups, imply that the human capital accumulation has virtually no marginal impact on growth for poor economies, whereas the variable is strongly significant outside of this group. Finally, we find some evidence of local convergence, as measured by the coefficient on initial income $\ln(Y/L)_i, 1960$. For the poor economies, the point estimate of -.791 is far higher than the single regime case and statistically significant at 5%. Further, the coefficient for the intermediate economies, -.172, is also negative, although not statistically significant. Interestingly, for the high income economies, the point estimate for the convergence variable is positive at .069, providing no evidence of convergence. This failure parallels the results of DeLong [1988] who rejected convergence over a much longer time span when studying economies with similar high initial incomes. These estimates imply that the gap between initial and steady state incomes has a half life of about 11 years for the poor economies, while for the intermediate economies the half life is over 8 times as long.

The second part of Table 4 reports the results of the associated constrained regressions. These results parallel the unconstrained results. We find that the intermediate economies have the largest physical capital share of .401, while the high income economies have a share of .333 (17% smaller) and the poor economies have a share of .306 (almost 25% smaller than the intermediate economies). In terms of the

Table 5
 Cross Section Regressions
 Literacy-Based Sample Breaks
 Dependent Variable: $\ln\left(\frac{Y}{L}\right)_{i,1985} - \ln\left(\frac{Y}{L}\right)_{i,1960}$

	$LR_{i,1960} < 50\%$	$50\% \leq LR_{i,1960} \leq 66\%$	$66\% < LR_{i,1960}$
Observations	48	12	36
Unconstrained Regressions			
constant	0.396 (1.73)	0.148 (0.416)	0.364 (0.945)
$\ln\left(\frac{Y}{L}\right)_{i,1960}$	-0.270 [†] (0.128)	-0.427 [†] (0.022)	-0.376 [†] (0.086)
$\ln\left(\frac{I}{Y}\right)_i$	0.324 [†] (0.114)	1.114 [†] (0.056)	0.686 [†] (0.209)
$\ln(n + g + \delta)_i$	-0.308 (0.486)	-0.113 (0.135)	-0.462 (0.414)
$\ln(SCHOO L)_i$	0.214 [†] (0.090)	0.164 [†] (0.048)	0.032 (0.223)
\bar{R}^2	0.28	0.97	0.29
σ_ϵ	0.34	0.08	0.33
Constrained Regressions			
θ	1.15 (1.38)	-12.6 [†] (1.54)	-1.56 (1.80)
α	0.294 [†] (0.094)	0.635 [†] (0.048)	0.595 [†] (0.145)
γ	0.266 [†] (0.059)	0.165 [†] (0.044)	0.047 (0.149)
\bar{R}^2	0.24	0.95	0.32
σ_ϵ	0.36	0.11	0.32

† denotes significance at asymptotic 5% level

human capital share, we find a point estimate near zero for the poor economies. In addition, the share for high income countries, .455, is 50% higher than the .302 intermediate economy estimate.

Estimating a three regime split based upon initial literacy, the quasi-likelihood function is maximized when countries are split into the categories $LR_{i,1960} < 50\%$, $50\% \leq LR_{i,1960} \leq 66\%$, and $66\% < LR_{i,1960}$. Table 5 presents the estimates of the constrained and unconstrained models.

As Table 5 indicates, segregation by the literacy rate also produces much heterogeneity across the unconstrained and constrained regressions. There is substantial variation in the ability of these regressions to explain the growth experiences of the countries in the different groups. The overall fit of the model is dramatically improved for the intermediate literacy economies. For the low and high literacy countries, the unconstrained and constrained regressions produce \bar{R}^2 estimates below .35 while for the intermediate group the estimates equal or exceed .95.

In terms of the unconstrained regressions, high and intermediate literacy economies exhibit far greater sensitivity to fluctuations in the physical investment ratio $\ln(I/Y)_i$ than low literacy countries. The estimated coefficients for the low, intermediate, and high income groups are .324, 1.114, and .686 respectively. All are statistically significant at the 5% level. Conversely, human capital investment, $\ln(SCHOO L)_i$, seems to have little effect on the high literacy economies, but does significantly affect the other groups. The estimated coefficients for the low, intermediate, and high income groups are .214, .164, and .032 respectively, with the latter being statistically insignificant at the 5% level. Perhaps most important, the estimated coefficients on $\ln(Y/L)_{i,1960}$ reveal a consistent pattern of local convergence within the literacy regimes. The low, intermediate, and high literacy coefficients, -.270, -.427, and -.376, are all negative and statistically significant. This pattern represents strong evidence, in light of the heterogeneity of the coefficient estimates, of local rather than global convergence in the Summers-Heston data.

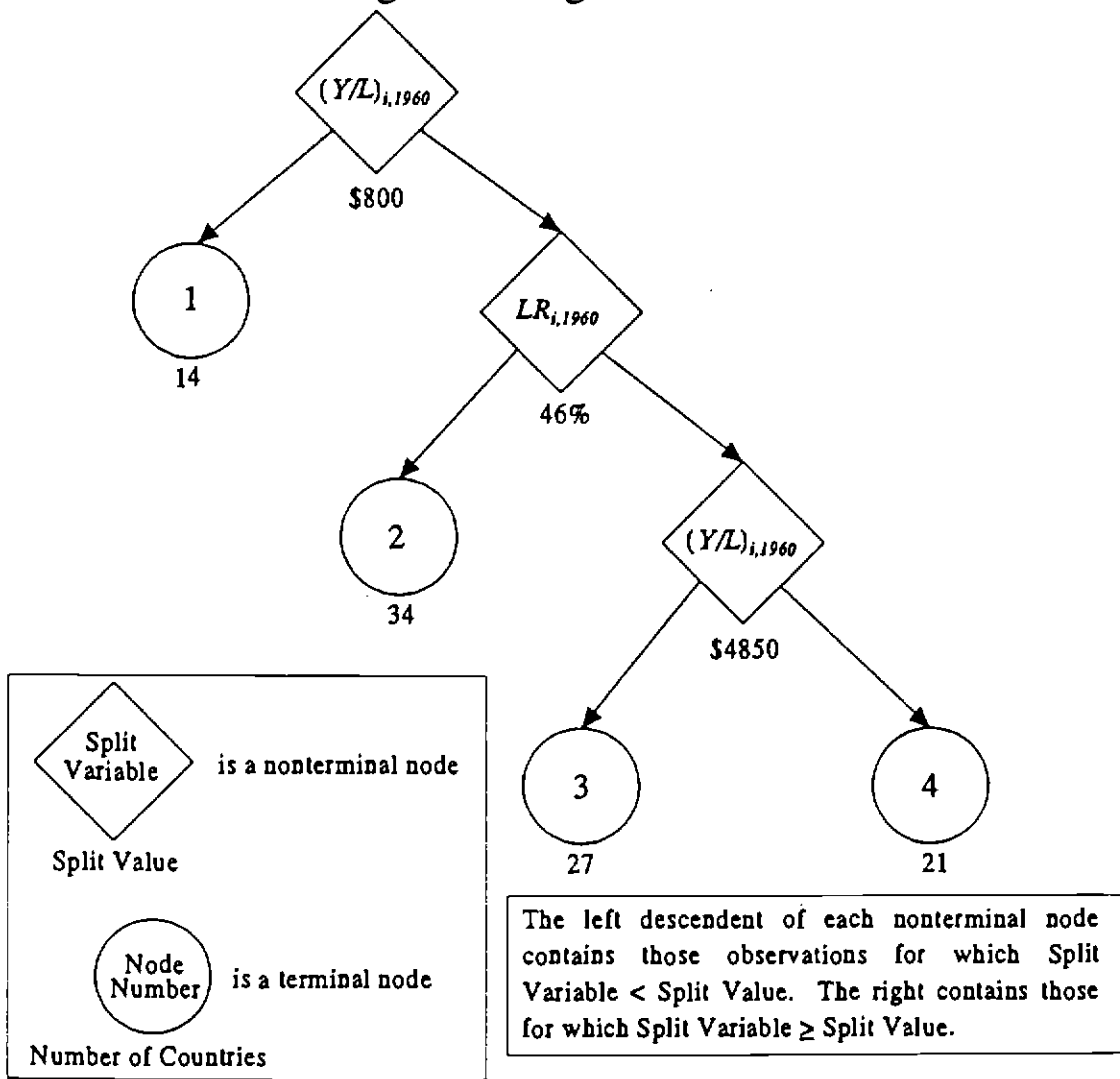
The constrained regression results in Table 5 reinforce the evidence of heterogeneity across groups. The physical capital share coefficient for low literacy countries, .294, is less than half the size of the .635 value for intermediate literacy economies, and the .595 estimate for high literacy economies. The human capital share declines as literacy increases with low to high literacy group estimates of .266, .165 and .047. In addition, the share is statistically significant only for the low and intermediate literacy economies. Given that our human capital variable measures secondary school enrollment, it is plausible that, for high literacy economies, the return to secondary education is low on the margin, while for relatively illiterate economies, large returns to secondary education still exist. This is difficult to reconcile, though, with the large human capital share for high initial income economies.

Using both income levels and literacy rates to identify different regimes, we have found substantial evidence of heterogeneity in production technologies and of local rather than global convergence in national economies. However, the characteristics of the regimes differ according to which variable is used to split the sample. These differences indicate the importance of exploring interactions between the control variables. At the same time, mechanically splitting the data to allow interactions will quickly eliminate all degrees of freedom; for example, a three-way split by income and literacy will create 9 categories for only 96 observations. Further, since the number of regimes and nature of the interactions are not dictated by any economic theory, it is desirable to employ a data sorting method which allows the data to endogenously select these features.

Regression tree estimates

We now turn to regression tree methods for identifying separate data regimes. This technique, described in Breiman *et al* [1984], provides a general nonparametric way of identifying multiple data regimes. The technique allows us to search for an unknown number of sample splits using more than one control variable.

Figure 1: Regression Tree



Constructing a regression tree is quite complicated; the specific algorithm is described in a Technical Appendix to the paper. Heuristically, the method works in two steps. First, all possible binary data splits based on either initial income and initial literacy are separately computed with the restriction that within each sample split there must be more countries than regressors. Equation (15) is estimated over the subsamples in each split and the total sum of squared residuals over all observations is computed. The split with the minimum sum of squared residuals is then taken as the first split of the data. Within each subsample of countries produced by the first split, this procedure is then repeated, i.e. all possible binary splits for either income or literacy are again computed. A second set of splits is added to the first by again choosing those splits which minimize the sum of squared residuals. This procedure is repeated until no more splits can be computed.

The set of sample splits at this point is certain to severely overestimate the number of regimes in the data since we have not accounted for the possibility of spurious splits due to the elimination of degrees of freedom in the regressions. The second step in the procedure "prunes" the tree by eliminating data splits which lead to (relatively) small reductions in the error variance. By eliminating data splits according to a penalty function which trades off error variance reduction against the number of splits, and by employing "cross-validation" methods which will provide unbiased estimates of residual variance, one can show that it is possible to consistently decompose the data set by regime.¹² The procedure bears some similarity to the use of information criteria to identify distributed lag lengths. No known asymptotic theory exists to test for the number of regimes in the data as uncovered by the regression tree. The virtue of the procedure lies in its ability to uncover multidimensional data splits.

The result of this procedure is the regression tree shown in Figure 1. Squares in

¹²The method is consistent in the sense that if the data exhibit a finite number of regimes, these regimes will be identified as the number of observations becomes large. Further, if all the data are generated by a single process, then the regression tree will asymptotically converge to one regime.

Table 6
Regression Tree Sample Breaks
Country Classification

Terminal Node Number			
1	2	3	4
Burkina Faso	Algeria	Madagascar	Austria
Burundi	Angola	South Africa	Belgium
Ethiopia	Benin	Hong Kong	Denmark
Malawi	Cameroon	Israel	Finland
Mali	Central African Rep.	Japan	France
Mauritania	Chad	Korea	Federal Republic of Germany
Niger	Congo, People's Rep.	Malaysia	Italy
Rwanda	Egypt	Philippines	The Netherlands
Sierra Leone	Ghana	Singapore	Norway
Tanzania	Ivory Coast	Sri Lanka	Sweden
Togo	Kenya	Thailand	Switzerland
Uganda	Liberia	Greece	United Kingdom
Zaire	Morocco	Ireland	Canada
Burma	Mozambique	Portugal	Trinidad and Tobago
	Nigeria	Spain	United States of America
	Senegal	Costa Rica	Argentina
	Somalia	Dominican Republic	Chile
	Sudan	El Salvador	Uruguay
	Tunisia	Jamaica	Venezuela
	Zambia	Mexico	Australia
	Zimbabwe	Nicaragua	New Zealand
	Bangladesh	Panama	
	India	Brasil	
	Jordan	Columbia	
	Nepal	Ecuador	
	Pakistan	Paraguay	
	Syria	Peru	
	Turkey		
	Guatemala		
	Haiti		
	Honduras		
	Bolivia		
	Indonesia		
	Papua New Guinea		

Table 7
 Cross Section Regressions
 Regression Tree Sample Breaks
 Dependent Variable: $\ln\left(\frac{Y}{L}\right)_{i,1985} - \ln\left(\frac{Y}{L}\right)_{i,1960}$

	Terminal Node Number			
	1	2	3	4
Observations	14	34	27	21
Unconstrained Regressions				
constant	3.46 (2.27)	-0.915 (1.79)	0.277 (1.42)	-7.26 [†] (1.59)
$\ln\left(\frac{Y}{L}\right)_{i,1960}$	-0.791 [†] (0.269)	-0.086 (0.131)	-0.316 [†] (0.123)	0.069 (0.139)
$\ln\left(\frac{I}{Y}\right)_i$	0.314 [†] (0.109)	0.129 (0.159)	1.110 [†] (0.165)	0.475 [†] (0.119)
$\ln(n + g + \delta)_i$	-0.429 (0.678)	-0.390 (0.489)	0.059 (0.451)	-1.75 [†] (0.270)
$\ln(SCHOOL)_i$	-0.028 (0.073)	0.469 [†] (0.095)	-0.114 (0.167)	0.341 [†] (0.141)
\bar{R}^2	0.57	0.52	0.57	0.82
σ_e	0.16	0.28	0.28	0.12
Constrained regressions				
θ	4.107 [†] (0.552)	0.539 (1.809)	-3.95 (2.67)	-11.0 (7.64)
α	0.306 [†] (0.083)	0.186 (0.123)	0.758 [†] (0.095)	0.333 [†] (0.100)
γ	-0.034 (0.083)	0.416 [†] (0.080)	-0.073 (0.114)	0.455 [†] (0.103)
\bar{R}^2	0.64	0.40	0.55	0.71
σ_e	0.19	0.32	0.30	0.18

† denotes significance at asymptotic 5% level

this figure indicate the splitting criteria for the sample; circles represent terminal nodes which contain different subsamples. The subsamples are: 1) $(Y/L)_{i,1960} < \$800$, 2) $\$800 \leq (Y/L)_{i,1960} \leq \4850 and $LR_{i,1960} < 46\%$, 3) $\$800 \leq (Y/L)_{i,1960} \leq \4850 and $46\% \leq LR_{i,1960}$, and 4) $\$4850 < (Y/L)_{i,1960}$. Note that the first and fourth subsamples contain precisely the countries classified as low and high income countries in Section 3. In effect, the regression tree has partitioned the intermediate income countries according to whether or not $LR_{i,1960}$ is greater or less than 46%. The fact that, given the opportunity to split the sample by either income or literacy, the regression tree shows a preference for income splits suggests that income dominates literacy as a variable useful in identifying separate regimes in the data.

Table 6 details the countries in each subsample. The Table indicates that there is substantial geographic homogeneity within each group. The low income/low literacy group is composed almost exclusively of poor African countries. The intermediate income/low literacy group is largely made up of relatively resource rich African economies and subcontinental Asian countries. Far eastern Asian and Latin American countries dominate the intermediate income/high literacy group. European economies make up most of the high income group. Any classification procedure such as a regression tree can lead to some individual countries being misassigned. For example, Japan is an obvious outlier in the third group. (We certainly would not want to claim that the aggregate production function for Japan is more similar to that of El Salvador than the United States!) The groupings do, however, seem quite reasonable overall and are certainly compatible with a local convergence interpretation.

Table 7 presents estimates of the unconstrained and constrained models for each of the subsamples. Perhaps the most striking feature of these estimates is how much they differ across subsamples. As we saw in Table 4, the estimated coefficient on $\ln(Y/L)_{i,1960}$ is significant for the first group and insignificant for the fourth group. The point estimates for the second and third subsamples, -0.086 and -0.316, are both negative although only the latter is significant. The regression tree has thus identified a group of

clearly locally convergent countries within the intermediate income countries.

Similar heterogeneity holds for other variables. The coefficient on $\ln(I/Y)_i$ is significant in the first, third, and fourth subsamples, but the subsample estimates vary greatly, ranging from .314 in the first subsample to 1.110 in the third subsample. The estimated coefficient on $\ln(SCHOOL)_i$ is insignificant for the first and third subsamples, and is over a third larger in the second subsample (0.469) than in the fourth (0.341).

Estimation of the constrained model produces vastly different estimates of both the physical and human capital shares across regimes. The estimated physical capital share in the third subsample (.758) is more than twice that in the first (.306) and fourth (.333) and is not statistically significant in the second (.186). The estimated human capital share are near zero for the first and third subsamples and are approximately equal for the second and fourth subsamples at .416 and .455. The fourth subsample is the only case where both shares are significant. Our estimates are strongly consistent with the view that different economies have access to different aggregate technologies.

The striking differences in the human capital share can be interpreted in different ways. One possibility is that economies go through production regimes which are indexed by different thresholds of human capital formation, in a way similar to the model of Azariadis and Drazen [1990]. Suppose that certain forms of organization of production within a firm or industry are constrained by the educational level of the labor force. Once these constraints no longer bind, then marginal increases in human capital would appear to have low marginal product, until an economy grows to the point where production is reorganized, creating a need for more human capital. In this case, the second and fourth nodes may represent regimes where human capital accumulation still augments the employed technology. Again, the different estimates might also reflect the weakness of the human capital variable, $\ln(SCHOOL)_i$. This variable only measures secondary school enrollment. If primary, secondary and college human capital formation have regime-specific output shares, then this variable may simply perform poorly in some cases.

Finally, it is interesting to note the pattern of labor shares across country groups: .728 for node 1, .398 for node 2, .315 for node 3, and .212 for node 4. These figures clearly illustrate how the labor share declines as an economy becomes more developed in terms of literacy and production. This path for the evolution of the aggregate production function suggests that the high productivity of advanced economies is due not only to capital deepening, but to the way in which capital per worker is converted into output per worker.¹³ The idea that high output economies more effectively utilize capital resources relative to low output economies is a common feature of many multiple equilibrium models, and is one way to interpret Romer's [1986] model of capital complementarities. The distinction between the differences in the capital/labor ratio and differences in the productivity of capital per worker is exploited by Durlauf and Johnson [1992] to identify the sources of cross-country income disparities.

5. Summary and conclusions

A large body of empirical work has concluded that international output data exhibit convergence when adequate account is taken of microeconomic heterogeneity. This literature is important as it bears upon the empirical relevance of various endogenous growth models. A major difficulty with these studies is that the natures of the null and alternative hypotheses are not made clear. In particular, the empirical convergence literature has assumed that a negative correlation between initial income and subsequent growth is necessarily evidence of convergence. This paper proposes a new set of empirical methods for studying convergence which explicitly allow for the possibility of multiple equilibria in the data.

Taking as a starting place the work of Mankiw, Romer and Weil [1990], we have

¹³Recall that for the two factor Cobb-Douglas technology, output per worker increases monotonically with the capital share.

reexamined the Summers-Heston data set to see whether the pattern of cross-country growth rates is compatible with a model of global convergence. We do this by explicitly allowing for the possibility that the data exhibit local rather than global convergence. Our approach uses information known at the start of the sample period to identify countries with similar initial conditions. We then see whether segregating countries into groups by initial conditions improves overall model fit. Conditioning initial income and/or initial literacy rates, specification tests support a multiple equilibrium interpretation. Using regression tree methods to find optimal splits in the data reveals substantial differences between the aggregate production functions of economies with different initial conditions. Consequently, our results demonstrate that the behavior of national growth rates in the postwar period is quite compatible with a multiple equilibrium perspective.

One important extension of our work is to see whether the apparent multiple regimes in the Heston-Summers data can be explicitly identified as arising from some of the production or demand complementarities which have been proposed as explanations for long run divergence. The identification of these complementarities is essential in understanding the policy implications of the endogenous growth literature.

Technical Appendix: Regression tree analysis

This Appendix contains a brief introduction to regression tree methods. We draw extensively from Breiman, *et al* [1984] throughout. The method is designed to uncover general forms of nonlinearity in data; Breiman *et al* show that the regression tree method is consistent in the sense that under suitable regularity conditions, estimates of the regression function converge to the true data generating process. The idea behind regression trees is that the prediction of y based on the vector X , which we call $d(X)$, may be improved (in the sense of reducing $E(y - d(X))^2$) by allowing the function $d(X)$ to change according to the values taken by some set of control variables Z .

A *tree*, $\{T, l, r\}$, is a finite nonempty set of positive integers, T , and two functions, $l(\cdot)$ and $r(\cdot)$, from T to $T \cup \{0\}$ such that, for each $t \in T$, (i) either $l(t) = r(t) = 0$ or $l(t) > t$ and $r(t) > t$, and (ii) if $t \neq \min\{t \in T\}$ there is exactly one $s \in T$, called the *parent* of t , such that, either, $t = l(s)$ or $t = r(s)$. The value of $\min\{t \in T\}$ is also called the *root* of the tree. Intuitively, these properties partially order the elements of T . In this ordering, each node t is either terminal or is followed by left and right elements $l(t)$ and $r(t)$.

Each element of T is called a *node* of the tree. Typically, $T = \{1, 2, \dots, \tau\}$ for some $\tau \geq 1$. For simplicity, we will abuse notation and use T to denote the tree $\{T, l, r\}$. The root node has no parent, and by (ii) above, every other node has a unique parent. Let $\text{parent}(\cdot)$ denote the function from T to $T \cup \{0\}$ defined so that $\text{parent}(\text{root}(T)) = 0$ and $\text{parent}(t)$ is the parent of t . The node t is an *ancestor* of the node s if t is the m -fold composition of $\text{parent}(s)$ for some m . If t is an ancestor of s then s is a *descendant* of t . When $l(t)$ and $r(t) \neq 0$, the nodes $l(t)$ and $r(t)$ are, respectively, the left and right descendants of t . A node is called *terminal* if $l(t) = r(t) = 0$, i.e. it has no descendants, otherwise it is called *nonterminal*. The set of terminal nodes is denoted \tilde{T} . The set of terminal descendants of t is denoted $\tilde{T}(t)$. Given $T^* \subset T$, define l^* and r^* from T^* to $T^* \cup \{0\}$ by

$$l^*(t) = \begin{cases} l(t) & \text{if } l(t) \in T^* \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad r^*(t) = \begin{cases} r(t) & \text{if } r(t) \in T^* \\ 0 & \text{otherwise} \end{cases} \quad (\text{A.1})$$

Then, T^* is a *subtree* of T if $\{T^*, l^*, r^*\}$ is a tree. If T^* is a subtree of T and $\text{root}(T^*) = \text{root}(T)$, then T^* is a *pruned subtree* of T , denoted $T^* \preceq T$. The relation \preceq is transitive and $T^* \prec T$ indicates $T^* \preceq T$ and $T^* \neq T$.

Let $Z \in \Omega \subseteq \mathbb{R}^n$ be the vector of control variables associated with X . The elements of Z may or may not be distinct from the variables in X . A *regression tree* partitions Ω by a sequence of binary splits, one at each nonterminal node. Let $\omega(t)$ be the subset of Ω associated with node t , so that $\omega(\text{root}(T)) = \Omega$ and $\bigcup_{t \in \tilde{T}} \omega(t) = \Omega$. For all nonterminal nodes, t , $\omega(l(t)) \cap \omega(r(t)) = \emptyset$ and $\omega(l(t)) \cup \omega(r(t)) = \omega(t)$. Then, for $Z \in \omega(t)$,

$$Z \in \begin{cases} \omega(l(t)) & \text{if } z_j < \gamma_j \in \mathbb{R} \\ \omega(r(t)) & \text{if } z_j \geq \gamma_j \in \mathbb{R} \end{cases} \quad (\text{A.2})$$

for some $1 \leq j \leq n$, where γ_j is the *split value* for z_j . For each $t \in \tilde{T}$, the predictor of y given $Z \in \omega(t)$ is $d_t(X)$.

There are two main steps to growing the sample regression tree from the sample $S = \{(y_i, X_i, Z_i), i = 1, \dots, N\}$.¹⁴ One first grows the largest tree allowed by the sample. Second, the tree is pruned in order to achieve an estimate of the optimal predictor $\hat{d}_t(X)$ for each $t \in \tilde{T}$. We restrict attention to the case where d_t is linear in its parameters and write $d_t(X) = \beta_t X$, where β_t is a vector conformable with X if the Z associated with X is in $\omega(t)$. Let $i(t) = \{1 \leq i \leq N \mid Z_i \in \omega(t)\}$ and define the within-node residual sum of squares $R(t) = \sum_{i \in i(t)} (y_i - \hat{d}_t(X_i))^2$ where $\hat{d}_t(X) = \hat{\beta}_t X$ and $\hat{\beta}_t$ is the within-node OLS estimator of β_t .¹⁵ The improvement in the sum of squares from a split at t is given by

¹⁴The application of the regression tree technique here has $y_i = \ln(Y/L)_{i,1985} - \ln(Y/L)_{i,1960}$, $X_i = [1, \ln(Y/L)_{i,1960}, \ln(I/Y)_i, \ln(n_i + g + \delta), \ln(\text{SCHOOL})_i]'$, and, $Z_i = [(Y/L)_{i,1960}, LR_{i,1960}]'$.

$\Delta_t = R(t) - R(l(t)) - R(r(t)) \geq 0$. Beginning at the root, we grow a tree by choosing, at each currently terminal node, t , the partition of $\omega(t)$ which maximizes Δ_t over choices of both z_j and γ_j , i.e. for each element of Z , we choose the split value which maximizes Δ_t and use that element of Z with the largest maximized Δ_t to perform the split. We continue in this way until there are too few observations in each terminal node to allow any further splits.

This tree is a very unparsimonious representation of the data and the final estimate of the error variance, $\sigma^2(T) = \frac{1}{N} \sum_{t \in \tilde{T}} R(t)$ (known as the "resubstitution" estimate), is an overly optimistic estimate of the model's accuracy as the model has been chosen to minimize this quantity. Further, no penalty has been imposed for the degrees of freedom lost as the tree was grown. The second step is to prune the tree. For any node, s , define the *error complexity measure*, $e_\alpha(s) = R(s) + \alpha$, where α is the *complexity parameter*, the cost of a terminal node. We then make a nonterminal node t terminal (i.e. eliminate all of its descendants) if $e_\alpha(t) \leq \sum_{s \in \tilde{T}(t)} e_\alpha(s)$. By letting α vary between 0 and ∞ , one induces a sequence of trees $T_1 \succ \dots \succ T_q$. T_1 is the largest tree that can be grown from the sample and T_q is the trivial tree containing only the root node. These are the candidates for the optimal tree.

The optimal tree is selected from this sequence using *cross-validation* estimates of the error variance of the model associated with each tree in the sequence. For a given tree, consider the observation (y_i, X_i) at node t . For each observation, a predictor $\hat{\beta}_t^{(i)} X_i$ is associated with each y_i , where $\hat{\beta}_t^{(i)}$ denotes the estimated coefficient vector at node t after omitting observation i . The prediction error for an observation will therefore equal $y_i - \hat{\beta}_t^{(i)} X_i$. The sample mean of $(y_i - \hat{\beta}_t^{(i)} X_i)^2$ over all observations is the cross-validation estimate of the error variance for the tree in question. Performing this procedure for each tree in the sequence generates a sequence of estimated error variances, one for each tree. The optimal tree is the one with the smallest estimated error variance.

¹⁵That is $\hat{\beta}_t = \left(\sum_{i(t)} X_i X_i' \right)^{-1} \sum_{i(t)} X_i y_i$.

Data Appendix

Number	Country	(Y/L) _{i,1960}	LR _{i,1960}	Growth Rate	Income Class	Literacy Class
1	Algeria	2485	10.0	4.8	I	L
2	Angola	1588	5.0*	0.8	I	L
3	Benin	1116	5.0*	2.2	I	L
4	Botswana	959	na	8.6	I	na
5	Burkina Faso	529	2.0*	2.9	L	L
6	Burundi	755	14.0*	1.2	L	L
7	Cameroon	889	19.0*	5.7	I	L
8	Central African Republic	838	7.0*	1.5	I	L
9	Chad	908	6.0	-0.9	I	L
10	People's Republic of the Congo	1009	16.0*	6.2	I	L
11	Egypt	907	26.0	6.0	I	L
12	Ethiopia	533	15.0+*	2.8	L	I
15	Ghana	1009	27.0*	1.0	I	L
17	Ivory Coast	1386	5.0*	5.1	I	L
18	Kenya	944	20.0*	4.8	I	L
20	Liberia	863	9.0*	3.3	I	L
21	Madagascar	1194	50.0†*	1.4	I	I
22	Malawi	455	25.0+*	4.8	L	L
23	Mali	737	2.0	2.1	L	L
24	Mauritania	777	5.0	3.3	L	L
25	Mauritius	1973	na	4.2	I	na
26	Morocco	1030	14.0	5.8	I	L
27	Mozambique	1420	8	1.4	I	L
28	Niger	539	1.0	4.4	L	L
29	Nigeria	1055	15.0*	2.8	I	L
30	Rwanda	460	16.0*	4.5	L	L
31	Senegal	1392	6.0*	2.5	I	L
32	Sierra Leone	511	7.0*	3.4	L	L
33	Somalia	901	2.0	1.8	I	L
34	South Africa	4768	57.0	3.9	I	I
35	Sudan	1254	13.0*	1.8	I	L
37	Tanzania	383	10.0	5.3	L	L
38	Togo	777	10.0	3.4	L	L
39	Tunisia	1623	16.0*	5.6	I	L
40	Uganda	601	35.0*	3.5	L	L
41	Zaire	594	31.0	0.9	L	L
42	Zambia	1410	29.0	2.1	I	L
43	Zimbabwe	1187	39.0*	5.1	I	L
46	Bangladesh	846	22.0*	4.0	I	L
47	Burma	517	60.0*	4.5	L	I
48	Hong Kong	3085	70.0	8.9	I	H
49	India	978	28.0*	3.6	I	L
52	Israel	4802	84.0*	5.9	I	H
53	Japan	3493	98.0*	6.8	I	H
54	Jordan	2183	32.0*	5.4	I	L
55	Republic of Korea	1285	71.0	7.9	I	H
57	Malaysia	2154	53.0	7.1	I	I
58	Nepal	833	9.0	2.6	I	L

60	Pakistan	1077	15.0*	5.8	I	L
61	Philippines	1668	72.0	4.5	I	H
63	Singapore	2793	75.0†*	9.2	I	H
64	Sri Lanka	1794	75.0*	3.7	I	H
65	Syrian Arab Republic	2382	30.0	6.7	I	L
67	Thailand	1308	68.0	6.7	I	H
70	Austria	5939	99.0	3.6	H	H
71	Belgium	6789	99.0†	3.5	H	H
73	Denmark	8551	99.0†	3.2	H	H
74	Finland	6527	99.0*	3.7	H	H
75	France	7215	99.0†	3.9	H	H
76	Federal Republic of Germany	7695	99.0†	3.3	H	H
77	Greece	2257	81.0	5.1	I	H
79	Ireland	4411	97.0*	3.8	I	H
80	Italy	4913	91.0*	3.8	I	H
83	Netherlands	7689	99.0†	3.6	H	H
84	Norway	7938	99.0†*	4.3	H	H
85	Portugal	2272	62.0	4.4	I	I
86	Spain	3766	87.0	4.9	I	H
87	Sweden	7802	99.0†*	3.1	H	H
88	Switzerland	10308	99.0†	2.5	H	H
89	Turkey	2274	38.0	5.2	I	L
90	United Kingdom	7634	99.0†	2.5	H	H
92	Canada	10286	99.0†	4.2	H	H
93	Costa Rica	3360	90.0†	4.7	I	H
94	Dominican Republic	1939	65.0	5.1	I	I
95	El Salvador	2042	49.0*	3.3	I	L
96	Guatemala	2481	32.0	3.9	I	L
97	Haiti	1096	15.0	1.8	I	L
98	Honduras	1430	45.0*	4.0	I	L
99	Jamaica	2729	82.0	2.1	I	H
100	Mexico	4229	65.0	5.5	I	I
101	Nicaragua	3195	57.0†	4.1	I	I
102	Panama	2423	73.0	5.9	I	H
103	Trinidad and Tobago	9253	93.0*	2.7	H	H
104	United States of America	12362	98.0*	3.2	H	H
105	Argentina	4852	91.0	2.1	H	H
106	Bolivia	1618	39.0	3.3	I	L
107	Brazil	1842	61.0	7.3	I	I
108	Chile	5189	84.0	2.6	H	H
109	Columbia	2672	63.0*	5.0	I	I
110	Ecuador	2198	68.0*	5.7	I	H
112	Paraguay	1951	75.0*	5.5	I	H
113	Peru	3310	61.0	3.5	I	I
115	Uruguay	5119	94.0*	0.9	H	H
116	Venezuela	10367	63.0*	1.9	H	I
117	Australia	8440	100.0†	3.8	H	H
119	Indonesia	879	39.0*	5.5	I	L
120	New Zealand	9523	99.0†	2.7	H	H
121	Papua New Guinea	1781	29.0	3.5	I	L

"Number" is the number given the country in the Summers and Heston [1988] data set.

na = not available.

† indicates that the Literacy Rate is for 1975 rather than 1960 as this is the next earliest available year.

* indicates that the Literacy Rate is for a year different, though no more than 2 years distant, from the specified year.

References

- Azariadis, Costas and Allan Drazen, "Threshold Externalities in Economic Development," *Quarterly Journal of Economics*, CV, 501-526, 1990.
- Barro, Robert J., "Economic Growth in a Cross Section of Countries," *Quarterly Journal of Economics*, CVI, 407-444, 1991.
- Baumol William J., "Productivity Growth, Convergence and Welfare: What the Long Run Data Show," *American Economic Review*, 76, 1072-1085, 1986.
- Bernard, Andrew B. and Steven N. Durlauf, "Interpreting Tests of the Convergence Hypothesis," Working Paper, Stanford University, 1991.
- Breiman, Leo, Jerome L. Friedman, Richard A. Olshen, and Charles J. Stone, *Classification and Regression Trees*, Wadsworth, Belmont Ca.: 1984.
- DeLong, J. Bradford, "Productivity Growth, Convergence and Welfare: Comment," *American Economic Review*, 78, 1138-1154, 1988.
- Durlauf, Steven N., "Nonergodic Economic Growth," NBER Working Paper no. 3719, 1991.
- Durlauf, Steven N. and Paul A. Johnson, "Technology Differences and International Income Disparities," Working Paper, Stanford University, in progress, 1992.
- Grossman, Gene M. and Elhanan Helpman, *Innovation and Growth in the Global Economy*, Cambridge: MIT Press, 1991.
- Lucas, Robert E., "On the Mechanics of Economic Development," *Journal of Monetary Economics*, 22, 3-42, 1988.
- Mankiw, N. Gregory, David Romer and David N. Weil, "A Contribution to the Empirics of Economic Growth," NBER Working Paper no. 3541, 1990.
- Murphy, Kevin, Andrei Shleifer, and Robert Vishny, "Industrialization and the Big Push," *Journal of Political Economy*, 97, 1003-1026, 1989.
- Quandt, Richard E., "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes," *Journal of the American Statistical Association*, 53, 873-880, 1958.
- Rauch, James E., "The Question of Per Capita Convergence in International

Consumption," UC San Diego Working Paper, 1989.

Romer, Paul M., "Increasing Returns and Long Run Growth," *Journal of Political Economy*, 94, 1002-1037, 1986.

Summers, Robert and Alan Heston, "A New Set of International Comparisons of Real Product and Prices: Estimates for 130 Countries," *Review of Income and Wealth*, 34, 1-26, 1988.

White, Halbert, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817-838, 1980.