JOURNAL OF APPLIED ECONOMETRICS, VOL. 10, 365–384 (1995)

MULTIPLE REGIMES AND CROSS-COUNTRY GROWTH BEHAVIOUR

STEVEN N. DURLAUF
Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706-1393, USA

PAUL A. JOHNSON
Department of Economics, Vassar College, Poughkeepsie, NY 12601, USA

SUMMARY

This paper provides some new evidence on the behaviour of cross-country growth rates. We reject the linear model commonly used to study cross-country growth behaviour in favour of a multiple regime alternative in which different economies obey different linear models when grouped according to initial conditions. Further, the marginal product of capital is shown to vary with the level of economic development. These results are consistent with growth models which exhibit multiple steady states. Our results call into question inferences that have been made in favour of the convergence hypothesis and further suggest that the explanatory power of the Solow growth model may be enhanced with a theory of aggregate production function differences.

1. INTRODUCTION

Starting with Baumol (1986), many authors have explored the behaviour of output growth across aggregate economies. Much of this research has been interested in determining whether economies exhibit convergence—defined as the tendency for per capita output differences due to initial conditions to disappear over time. The convergence question has important implications for growth theory. Convergence holds in the classic Solow growth model given a concave aggregate production function. In contrast, the new growth theory pioneered by Romer (1986) and Lucas (1988) shows how increasing returns to scale can cause differences in initial conditions to persist.

Much of the empirical work on convergence has been concerned with identifying a negative coefficient in the regression of a country's growth over a fixed period on its initial output, so that poorer countries grow faster on average than richer ones. As initially shown in Barro (1991) this negative coefficient may be found for a large cross-section of countries when one controls for factors such as education, investment rates, and political stability. Perhaps the most persuasive evidence in this regard is due to Mankiw *et al.* (1992), who study a regression in which the control variables are those directly suggested by a human capital-augmented version of the Solow model. These authors find not only strong evidence of convergence but that the regression fulfils the cross-coefficient restrictions imposed by the Solow model.

One important assumption underlying the bulk of cross-country growth studies is that all countries obey a common linear specification. While a linear specification holds under the Solow model, it does not hold for some new growth alternatives. A class of growth models, starting with Azariadis and Drazen (1990), produces multiple locally stable steady states in per capita output. Cross-country growth behaviour in these models is typically nonlinear, exhibiting

CCC 0883–7252/95/040365–20

Received March 1993
Revised December 1994

multiple regimes as countries associated with the same steady state obey a common linear regression. Conventional convergence tests will have difficulty distinguishing between these multiple steady-state models and the Solow model. As shown in Bernard and Durlauf (1993), a linear regression applied to data generated by economies converging to multiple steady states can produce a negative initial income coefficient. Intuitively, the initial income coefficient in the misspecified linear model inherits the convergence exhibited among countries associated with a common steady state in the correctly specified multiple regime growth process.

This paper re-examines the Summers–Heston data set in order to identify whether multiple regimes in cross-country growth behaviour are present. This exercise is of interest both from the perspective of better understanding the statistical properties of the data set results as well as evaluating the compatibility of cross-country growth patterns with multiple steady-state models. Our conclusions are twofold. First, we reject the null hypothesis that all countries obey a common linear model. This means that the equation estimated by previous authors to show the presence of convergence is misspecified. Second, by using regression tree analysis to identify countries obeying a common linear model, we find subsets of countries which appear to possess very different production functions. These differences in turn suggest that more developed countries have higher output–labour ratios than implied by their capital–labour ratios alone.

Section 2 reviews the link between some growth models and cross-section regressions. Section 3 describes the data we analyse. Section 4 performs specification tests on cross-country regressions. Using initial output and literacy rates to segregate countries, a single regime specification is rejected. Section 5 uses regression tree techniques to identify groups of countries obeying a common linear model. Section 6 provides some caveats to the interpretation of our results. Section 7 contains the conclusions. Data and technical appendices follow.

2. CONVERGENCE AND CROSS-SECTION BEHAVIOUR

To derive the cross-country implications of the Solow growth model, we follow the analysis in Mankiw *et al.* (1992), which we subsequently denote as M–R–W, and consider the case where aggregate output in country i at t, $Y_{i,t}$, is determined by a Cobb–Douglas production function taking as arguments the level of technology A_t, labour input $L_{i,t}$, physical capital input $K_{i,t}$, and human capital input $H_{i,t}$:

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

All variables are assumed to evolve in continuous time. The level of technology and labour grow at constant rates g and n_i, respectively. Each country augments its physical and human capital stocks at the constant savings rates s_i^k and s_i^h while both stocks depreciate at the same rate δ. This induces capital accumulation equations of the form $\mathrm{d}K_{i,t}/\mathrm{d}t = s_i^k Y_{i,t} - \delta K_{i,t}$ and $\mathrm{d}H_{i,t}/\mathrm{d}t = s_i^h Y_{i,t} - \delta H_{i,t}$. As a result, over any interval T to $T+\tau$ output per worker $(Y/L)_{i,t}$, obeys

$$\ln(Y/L)_{i,T+\tau} - \ln(Y/L)_{i,T} = g\tau + (1 - \mathrm{e}^{-\lambda_i \tau}) \times \left(\Theta + \frac{\alpha}{1-\alpha-\gamma} \ln(s_i^k) + \frac{\gamma}{1-\alpha-\gamma} \ln(s_i^h) - \frac{\alpha+\gamma}{1-\alpha-\gamma} \ln(n_i + g + \delta) - \ln(Y/L)_{i,T} \right) \tag{2}$$

Here $\Theta = 1/(1-\alpha-\gamma)\ln(\phi) - \ln(A_0) - gT$ and $\lambda_i = (1-\alpha-\gamma)(n_i + g + \delta)$; λ_i is the country-specific convergence rate towards the steady state.

Equation (2) explains cross-country growth rates as the result of a common technology which is concave in the two capital stocks and country-specific input growth. The equation places nonlinear restrictions across the regression coefficients and is typically referred to as the 'constrained' version of the Solow model. Relaxation of these restrictions while assuming that $\lambda_i = \lambda \ \forall \ i$, as M–R–W do, produces the 'unconstrained' law of motion estimated by Barro and others:

$$\ln(Y/L)_{i,T+\tau} - \ln(Y/L)_{i,T} = \zeta + \beta \ln(Y/L)_{i,T} + \Pi \mathbf{X}_i + \varepsilon_i, \qquad i = 1, \ldots, N \tag{3}$$

where $\mathbf{X}_i = (\ln(s_i^k), \ln(s_i^h), \ln(n_i + g + \delta))$.

A negative value for $-(1 - \mathrm{e}^{-\lambda_i \tau})(\alpha + \gamma)/(1 - \alpha - \gamma)$ in the constrained regression or for β in the unconstrained regression has been taken as evidence of convergence, corresponding to the intuition that convergence occurs when low per capita output economies grow more quickly than high per capita output ones.

A number of new growth models are based on the idea that there exists a range of human or physical capital levels over which the aggregate production function is not concave, which will lead to different long-run steady states for different initial conditions. For example, Azariadis and Drazen (1990) argue that there exist human or physical capital accumulation thresholds which induce shifts in aggregate productivity.[1] One version of the Azariadis–Drazen model replaces equation (1) with a production function embodying a physical capital threshold $\tilde{K}(t)$, and a human capital threshold $\tilde{H}(t)$, which may depend on time,[2] such that

$$Y_{i,t} = \phi (K_{i,t})^{\alpha_j} (H_{i,t})^{\gamma_j} (A_t L_{i,t})^{1-\alpha_j-\gamma_j} \tag{4}$$

where

$$\alpha_j = \alpha_1 \text{ if } K_{i,t} < \tilde{K}(t),\ \alpha_2 \text{ otherwise};\ \gamma_j = \gamma_1 \text{ if } H_{i,t} < \tilde{H}(t),\ \gamma_2 \text{ otherwise} \tag{5}$$

This type of non-convexity will, for some values of the thresholds $\tilde{H}(t)$ and $\tilde{K}(t)$, generate multiple steady-state equilibria with an associated cross-section law of motion:

$$\ln(Y/L)_{i,T+\tau} - \ln(Y/L)_{i,T} = g\tau + (1 - \mathrm{e}^{-\lambda_i \tau})$$
$$\times \left(\Theta_j + \frac{\alpha_j}{1-\alpha_j-\gamma_j} \ln(s_i^k) + \frac{\gamma_j}{1-\alpha_j-\gamma_j} \ln(s_i^h) - \frac{\alpha_j+\gamma_j}{1-\alpha_j-\gamma_j} \ln(n_i + g + \delta) - \ln(Y/L)_{i,T} \right) \tag{6}$$

where $\lambda_i = (1 - \alpha_j - \gamma_j)(n_i + g + \delta)$, and $\Theta_j = 1/(1 - \alpha_j - \gamma_j)\ln(\phi) - \ln(A_0) - gT$, with α_j and γ_j determined by equation (5).

A country possessing this technology will follow one of four distinct Solow-type laws of motion, depending on the relationship between $(K_{i,t}, H_{i,t})$ and $(\tilde{K}(t), \tilde{H}(t))$. As a result, the cross-section regressions (2) and (3) are correctly specified for subsets of countries when the aggregate technology obeys (4) and (5). The testable implications of (1) versus (4) and (5) are summarized in the requirement that the cross-section data are generated by a common linear law of motion whose coefficients obey the constraints embedded in (2). Conversely, Bernard and Durlauf (1993) show how a negative initial output coefficient can occur when (2) or (3) is estimated from data generated by (4) and (5), so that the standard convergence test cannot differentiate between the two models.[3]

This example motivates our empirical strategy for analysing cross-country growth behaviour

[1] Models such as those of Murphy *et al.* (1989) and Durlauf (1993) in which economies exhibit multiple steady states due to coordination failure produce similar multiple regime implications for the data.

[2] We allow for these thresholds to depend on time in order to account for factors such as technical change or population growth which might produce steady-state capital growth without causing a country to undergo the sort of industrial transformation envisioned by models such as that of Azariadis and Drazen (1990).

[3] Quah (1993) provides a related critique of the interpretability of the initial output coefficient.

by determining whether the data obey a single Solow-type growth equation or whether the data exhibit multiple regimes in the sense that subgroups of countries identified by initial conditions obey distinct Solow-type regressions.

3. DATA

All cross-country growth rates we employ are based upon the Summers–Heston (1988) international output estimates; the sample of countries and a number of associated data series are contained in the data appendix.[4] The variables are:

$(Y/L)_{i,t}$ = 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 1960–85.
n_i = growth rate of the working-age population, country i, annual average 1960–85.
$SCHOOL_i$ = fraction of the working-age population enrolled in secondary school, country i, annual average 1960–85.
$LR_{i,1960}$ = 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 $g = 0{\cdot}02$ (implying that $g\tau = 0{\cdot}5$, a value that we impose in estimation) and $\delta = 0{\cdot}03$, 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.

4. SPECIFICATION TESTS FOR MULTIPLE REGIMES

In this section we attempt to identify the presence of multiple 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

$$\ln(Y/L)_{i,1985} - \ln(Y/L)_{i,1960} = \zeta + \beta \ln(Y/L)_{i,1960} + \pi_1 \ln(I/Y)_i + \pi_2 \ln(n_i + g + \delta) + \pi_3 \ln(SCHOOL)_i + \varepsilon_i \quad (7)$$

by least squares over each subgroup. This estimate represents the unconstrained version of the Solow model. We separately estimate a constrained version of the model by imposing cross-coefficient restrictions defined by (2).

We consider two different control variables to group countries with initial conditions.[6] The first variable we employ is per capita output at the beginning of the sample period, $(Y/L)_{i,1960}$. Most models of multiple steady states predict that if economies are concentrated around several steady states, then their initial per capita output levels will fall into non-overlapping categories. Second,

[4] Besides Summers and Heston (1988), the primary sources for the data set are the World Bank's *World Tables* and *World Development Report*.
[5] For 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 much in a two-year period so the resultant error is probably small. Two countries studied by M–R–W, Botswana and Mauritius, are omitted due to lack of data on literacy.
[6] The use of split variables which are known at the beginning of the sample under study is necessary to avoid the selection bias problem noted by DeLong (1988).

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 differences in the level of social and economic development rather than the current level of economic activity.[7] Alternatively, these variables may be interpreted as proxies for identifying threshold effects associated with the unobserved physical and human capital stocks.

Table I 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.[8] The first panel of the table divides countries into two equal groups by segregating high and low initial output 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 using each of these variables. The third panel allows interactions between the variables. In this case, we divide the countries according to whether they lie in the high or low half of the sample according to the two controls. This segregation results in four categories: high-output/high-literacy (42 countries), high-output/low-literacy (6 countries), low-output/high-literacy (6 countries) and low-output/low-literacy (42 countries).[9]

As Table I indicates, there is substantial evidence that the laws of motion for growth within each subgroup are different. For three of the four initial output 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%.[10] Further, we strongly reject coefficient equality across splits for both the unconstrained and constrained regressions in the interactive four-regime specification. This change in the significance level indicates the importance of both variables in identifying data regimes.

Table II reports the original M–R–W regression along with estimates of the regressions

Table I. Specification tests for different regimes

Subsamples defined by	Unconstrained regressions	Constrained regressions
Two-way split based on		
$(Y/L)_{i,1960}$	0·009	0·218
$LR_{i,1960}$	0·011	0·112
Three-way split based on		
$(Y/L)_{i,1960}$	0·029	0·011
$LR_{i,1960}$	0·404	0·000
Four-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 the null hypothesis that the parameters of the indicated models are constant across the indicated subsamples. Splits are described in the text.

[7] This distinction is also relevant for coordination-based models with multiple regimes.

[8] Following Barro (1991) and others, we employ heteroscedasticity-corrected test statistics and standard error estimates based on White (1980), in order to permit error variances to differ across countries. White's heteroscedasticity test reveals some evidence against a homoscedastic null. Assuming homoscedasticity in the calculation of the Wald statistics increases the number of rejections of the single regime model.

[9] The two-way output 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 $72\% < LR_{i,1960}$. The data appendix records the three-way splits for various countries by identifying each as falling into a high (H), intermediate (I) or low (L) output or literacy class.

[10] See Rauch (1989) for corroborating evidence of literacy-based regime differences.

Table II. Cross-section regressions: initial output and literacy-based sample breaks: dependent variable: $\ln(Y/L)_{i,1985} - \ln(Y/L)_{i,1960}$

	M–R–W	$(Y/L)_{i,1960} < 1950$ and $LR_{i,1960} < 54\%$	$1950 \leq (Y/L)_{i,1960}$ and $54\% \leq LR_{i,1960}$
Observations	98	42	42
		Unconstrained regressions	
Constant	3·04[a]	1·40	0·450
	(0·831)	(1·85)	(0·723)
$\ln(Y/L)_{i,1960}$	−0·289[a]	−0·444[a]	−0·434[a]
	(0·062)	(0·157)	(0·085)
$\ln(I/Y)_i$	0·524[a]	0·310[a]	0·689[a]
	(0·087)	(0·114)	(0·170)
$\ln(n+g+\delta)_i$	−0·505	−0·379	−0·545
	(0·288)	(0·468)	(0·283)
$\ln(SCHOOL)_i$	0·233[a]	0·209[a]	0·114
	(0·060)	(0·094)	(0·164)
$\bar{R}^2$	0·46	0·27	0·48
σ_ε	0·33	0·34	0·30
		Constrained regressions	
Θ	−2·56[a,b]	2·29	−0·395
	(1·14)	(1·17)	(1·24)
α	0·431[a]	0·275[a]	0·509[a]
	(0·061)	(0·097)	(0·098)
γ	0·241[a]	0·217[a]	0·108
	(0·046)	(0·061)	(0·094)
$\bar{R}^2$	0·42	0·28	0·50
σ_ε	0·34	0·34	0·29

[a] Significance at asymptotic 5% level.
[b] This equation has been reestimated under the restriction $\lambda_i = (1-\alpha-\gamma)(n_i+g+\delta)$, 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); $\alpha = 0{\cdot}48$ (0·07); $\gamma = 0{\cdot}23$ (0·05); $\bar{R}^2 = 0{\cdot}46$; and $\sigma_\varepsilon = 0{\cdot}33$.

associated with the high initial output/high initial literacy and low initial output/low initial literacy splits described above. (The high initial output/low initial literacy and low initial output/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 output, $\ln(Y/L)_{i,1960}$, is approximately equal for the high-output/high-literacy and low-output/low-literacy groups at −0·434 and −0·444, respectively; these estimates are much larger than the −0·289 estimate for the whole sample. This difference reveals a faster convergence rate for the subsamples than for the single regime. Further, the $\ln(I/Y)_i$ coefficient for high-output/high-literacy countries is 0·689, which is over twice as large as the 0·310 estimate for the low-output/low-literacy countries and over 25% higher than the 0·524 estimate for the whole sample. Similarly, the implied physical capital share in output for the constrained regressions, α, is far larger for the high-output/high-literacy countries at 0·509 than for the low-literacy/low-output countries at 0·275, and somewhat larger than the 0·431 share for the whole sample. Conversely, the low-output/low-literacy countries exhibit a much larger coefficient for the human capital investment measure $\ln(SCHOOL)_i$ as well as the associated human capital output share γ than high-output/

high-literacy countries, although both subsample estimates are below those for the whole sample. These estimates suggest that the aggregate production function differs substantially across subsamples.

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 $\mathbf{X}_i$ 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-regime model is robust to the addition of Barro's variables to those dictated by the strict Solow model. The variables we used are:[11]

$AFRICA_i$ = 1 if country i is in sub-Saharan Africa
$ASSASS_i$ = number of assassinations per million population per year, country i
$(G^c/Y)_i$ = average ratio of government consumption (exclusive of defence and education) to GDP, country i
$LATAMER_i$ = 1 if country i is in Latin America
$LIT60_i$ = adult literacy rate in 1960, country i[12]
$MIXED_i$ = 1 if country i has a mixed free enterprise/socialistic economic system
$PPI60DEV_i$ = deviation from sample mean of the 1960 purchasing power parity value for the investment deflator, country i
$PRIM60_i$ = primary-school enrollment rate, country i, 1960
REV_i = number of revolutions and coups per year, country i
$SEC60_i$ = secondary-school enrollment rate, country i, 1960[13]
SOC_i = 1 if country i has a socialist economic system
$STTEAPRI_i$ = 1960 primary school student–teacher ratio, country i
$STTEASEC_i$ = 1960 secondary school student–teacher ratio, country i.

We focus on the differences between the low-output/low-literacy and the high-output/high-literacy groups of countries identified above.[14] Specifically, we re-estimate equations (2) and (3) for these groups after adding subsets of the Barro variables corresponding to the different combinations of regressors reported in Barro (1991), again testing the significance of the differences in the estimated coefficients between groups.

Table III gives the results. The first line of the table establishes the significance of the differences between the two groups with no additional control variables. The remainder of the table shows the results when the controls are the regressors from equations 1 and 11 to 14 in Barro's Table 1. In no case does the marginal significance value exceed 0·002. The evidence of multiple regimes thus appears to be robust to the addition of these additional control variables.[15]

[11] All the data are from Barro and Wolf (1989).

[12] This variable is measured differently from the variable $LR_{i,1960}$ that we use to split up the sample but for the 94 countries for which there are data on both the correlation coefficient between them is 0·96.

[13] $SEC60_i$ differs from $SCHOOL_i$ as it measures the ratio of secondary students to the population between 12 and 17 rather than to all working age persons and because it equals a point estimate for 1960 rather than an average over 1960–85.

[14] We do not consider the high-output/low-literacy and low-output/high-literacy groups because of the small number of observations in each.

[15] We omit the variables in the subsequent analysis in order to highlight the differences between the single and multiple regime versions of the Solow growth model, since (1) inclusion of the variables has no qualitative effect on the results and (2) these control variables are *ad hoc* additions to the standard Solow model.

Table III. Specification tests: robustness check

Additional control variables	Marginal significance value
None	0·001
Equation (1)	0·002
Equation (11)	0·000
Equation (12)	0·000
Equation (13)	0·000
Equation (14)	0·000

This table shows the marginal significance levels for the Wald tests of null hypothesis that the coefficients on the variables dictated by the Solow model are the same in the low-output/low-literacy and high-output/high-literacy groups described in the text when the indicated sets of Barro control variables are added to the model. The sets of controls are:

Equation (1): $SEC60_i$, $PRIM60_i$, $(G^c/Y)_i$, REV_i, $ASSASS_i$, $PPI60DEV_i$
Equation (11): $STTEAPRI_i$, $STTEASEC_i$, $(G^c/Y)_i$, REV_i, $ASSASS_i$, $PPI60DEV_i$
Equation (12): $SEC60_i$, $PRIM60_i$, $LIT60_i$, $(G^c/Y)_i$, REV_i, $ASSASS_i$, $PPI60DEV_i$
Equation (13): $SEC60_i$, $PRIM60_i$, $(G^c/Y)_i$, REV_i, $ASSASS_i$, $PPI60DEV_i$, SOC_i, $MIXED_i$
Equation (14): $SEC60_i$, $PRIM60_i$, $(G^c/Y)_i$, REV_i, $ASSASS_i$, $PPI60DEV_i$, $AFRICA_i$, $LATAMER_i$

5. REGRESSION TREE ESTIMATES OF COUNTRY GROUPS

Although the exogenously imposed data splits of the previous section permit straightforward specification testing, they do not address the problem of identifying economies with common laws of motion. In order to identify economies whose growth behaviour obeys a common statistical model, it is necessary to allow the data to determine the location of the different regimes. At the same time, mechanically splitting the data by initial conditions in order to produce multiple regimes will quickly eliminate all degrees of freedom; for example, a three-way split by output and literacy creates nine categories for only 96 observations. Further, economic theory provides no prior guidance as to either the number of regimes or the way in which the different variables defining initial conditions interact in determining regimes. Therefore, it is desirable to employ a data-sorting method which allows the data to select these features endogenously.

Regime identification is performed based on regression tree analysis.[16] This technique, described in Breiman *et al.* (1984), provides a general non-parametric way of identifying multiple data regimes from a set of control variables. The technique allows one to search for an unknown number of sample splits using multiple control variables. Intuitively, the procedure approximates the growth process as a union of piecewise linear functions, where observations are grouped by initial conditions. The actual sorting algorithm is quite complicated and is described in the technical appendix. No asymptotic theory exists to test the statistical significance of the number of regimes 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 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) $\$4850 < (Y/L)_{i,1960}$.

[16] We have also examined endogenous data splitting in which the number of different regimes is set *a priori* with break points chosen to maximize a quasi-log likelihood function. Setting the number of regimes at three and using each of initial output or initial literacy to order the data produced statistically significant splits.

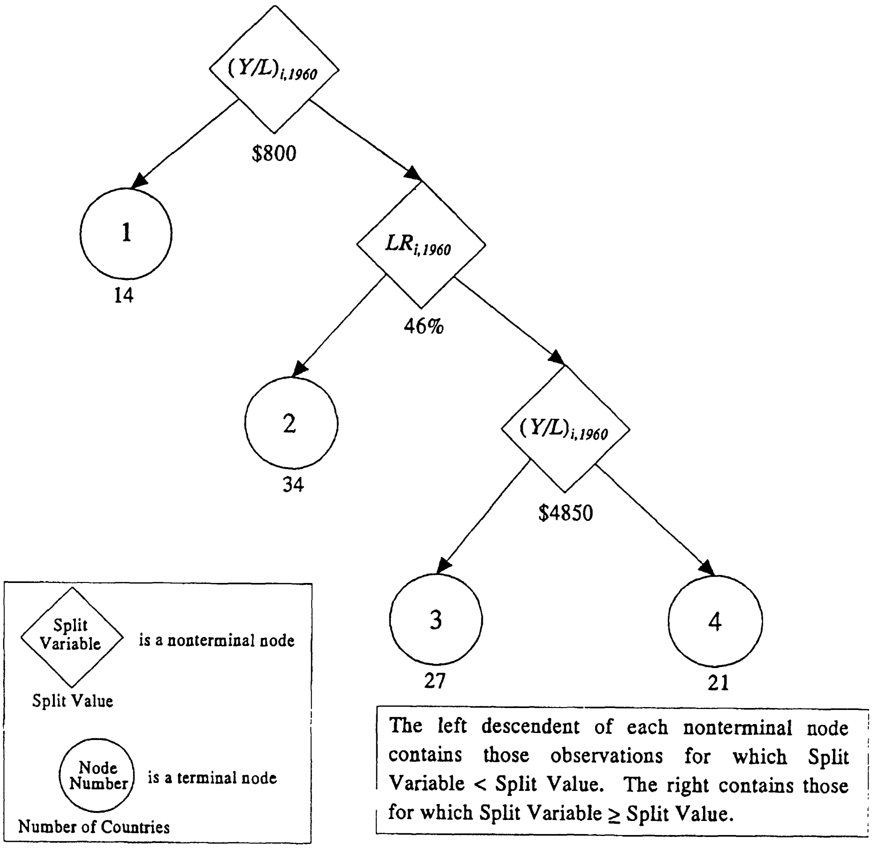

Figure 1. Regression tree

The regression tree partitions the sample into low-, intermediate-, and high-output countries and then further partitions the intermediate-output countries into low- and high-literacy countries. The fact that, given the opportunity to split the sample by either output or literacy, the regression tree shows a preference for output splits suggests that output dominates literacy as a variable useful in identifying multiple regimes in the data.

Table IV details the countries in each subsample. The table indicates that there is substantial geographic homogeneity within each group. The low-output/low-literacy group is composed almost exclusively of poor African countries. The intermediate-output/low-literacy group is largely made up of relatively resource-rich African countries and subcontinental Asian countries. Far Eastern Asian and Latin American countries dominate the intermediate-output/high-literacy group. North American and European countries dominate the high-output group.[17]

Table V presents estimates for each of the subsamples. In terms of overall fit, we find some improvement over the single-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

[17] We performed several specification tests on the regression tree splits to see if there is any residual nonlinearity. Tests based on the addition of quadratic or cubic terms to the within-group linear regressions or based on additional data splits as determined by initial conditions produced little evidence of within-group nonlinearity.

Table IV. 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	Brazil	
	Jordan	Columbia	
	Nepal	Ecuador	
	Pakistan	Paraguay	
	Syria	Peru	
	Turkey		
	Guatemala		
	Haiti		
	Honduras		
	Bolivia		
	Indonesia		
	Papua New Guinea		

poorest economies, we explain 57%, for intermediate-output economies with low literacy rates 52%, for the intermediate-output economies with high literacy rates 57%, and for high-output economies fully 82% of the total growth variation. Similar results hold for the constrained regressions.

Perhaps the most striking feature of these estimates is how much they differ across subsamples. For example, the estimated coefficient on $\ln(Y/L)_{i,1960}$ is $-0{\cdot}791$ and significant for the first group while it is 0·069 but insignificant for the fourth group. This failure to find evidence of convergence among the high-output economies parallels the results of DeLong (1988) who rejected convergence over a much longer time span when studying economies with similar high initial outputs. The point estimates for the second and third subsamples, $-0{\cdot}086$ and $-0{\cdot}316$, are both negative although only the latter is significant. The regression tree thus identifies a convergent subgroup within the intermediate-output countries.

Table V. Cross-section regressions: regression tree sample breaks: dependent variable: $\ln(Y/L)_{i,1985} - \ln(Y/L)_{i,1960}$

	Terminal node number			
	1	2	3	4
Observations	14	34	27	21
		Unconstrained regressions		
Constant	3·46	−0·915	0·277	−7·26[a]
	(2·27)	(1·79)	(1·42)	(1·59)
$\ln(Y/L)_{i,1960}$	−0·791[a]	−0·086	−0·316[a]	0·069
	(0·269)	(0·131)	(0·123)	(0·139)
$\ln(I/Y)_i$	0·314[a]	0·129	1·110[a]	0·475[a]
	(0·109)	(0·159)	(0·165)	(0·119)
$\ln(n+g+\delta)_i$	−0·429	−0·390	0·059	−1·75[a]
	(0·678)	(0·489)	(0·451)	(0·270)
$\ln(SCHOOL)_i$	−0·028	0·469[a]	−0·114	0·341[a]
	(0·073)	(0·095)	(0·167)	(0·141)
$\bar{R}^2$	0·57	0·52	0·57	0·82
σ_ε	0·16	0·28	0·28	0·12
		Constrained regressions		
Θ	4·107[a]	0·539	−3·95	−11·0
	(0·552)	(1·809)	(2·67)	(7·64)
α	0·306[a]	0·186	0·758[a]	0·333[a]
	(0·083)	(0·123)	(0·095)	(0·100)
γ	−0·034	0·416[a]	−0·073	0·455[a]
	(0·083)	(0·080)	(0·114)	(0·103)
$\bar{R}^2$	0·64	0·40	0·55	0·71
σ_ε	0·19	0·32	0·30	0·18

[a] Significance at asymptotic 5% level.

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 0·314 in the first subsample to 1·110 in the third. 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 (0·758) is more than twice that in the first (0·306) and fourth (0·333) and is not statistically significant in the second (0·186). The estimated human capital shares are near zero for the first and third subsamples and are approximately equal for the second and fourth subsamples at 0·416 and 0·455. 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 that 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 labour force. Once these constraints no longer bind, then marginal increases in human capital will 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 augments the utilized technology. Alternatively, 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. In general, if the magnitude of measurement error for any of the right-hand side variables correlates with the initial conditions we use for sample splitting, spurious production function differences could be identified.

Finally, it is interesting to compute the pattern of labour shares across country groups: 0·728 for node 1, 0·398 for node 2, 0·315 for node 3, and 0·212 for node 4. These figures illustrate how the labour 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 also to the way in which capital per worker is converted into output per worker.[18] The idea that high-output economies more effectively utilize capital relative to low-output economies is a feature of many multiple steady-state models, and is one way to interpret either the Azariadis and Drazen (1990) model of threshold externalities or the Durlauf (1993) model of local technological spillovers, and is consistent with the finding in Dowrick and Gemmell (1991) of capital productivity differences between rich and poor countries.[19]

6. CAVEATS

We raise three caveats in the interpretation of our results.

6.1. Identification

While our results illustrate how the standard cross-section growth regression is misspecified and how a nonlinear generalization of the standard regression exists which is compatible with a multiple steady-state alternative, it is important to emphasize a sense in which the presence or absence of convergence is not identified by the analysis. Simply put, the contrasting behaviour of economies with different initial conditions is compatible both with a model in which economies pass through distinct phases of development towards a unique steady state as well as one in which multiple steady states exist. This basic identification problem in interpreting the long-run implications of multiple growth regimes is illustrated in Figure 2, where a single capital type is assumed. If the production function follows the solid line for all capital/labour ratios k, then developed and underdeveloped economies will fail to converge. Alternatively, if the broken line represents the aggregate production function for capital/labour ratios between the capital levels of the underdeveloped economies and k^{T}, then convergence holds.

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

[19] At the same time, the pattern of human and physical capital coefficients is not a specific implication of any growth model of which we are aware.

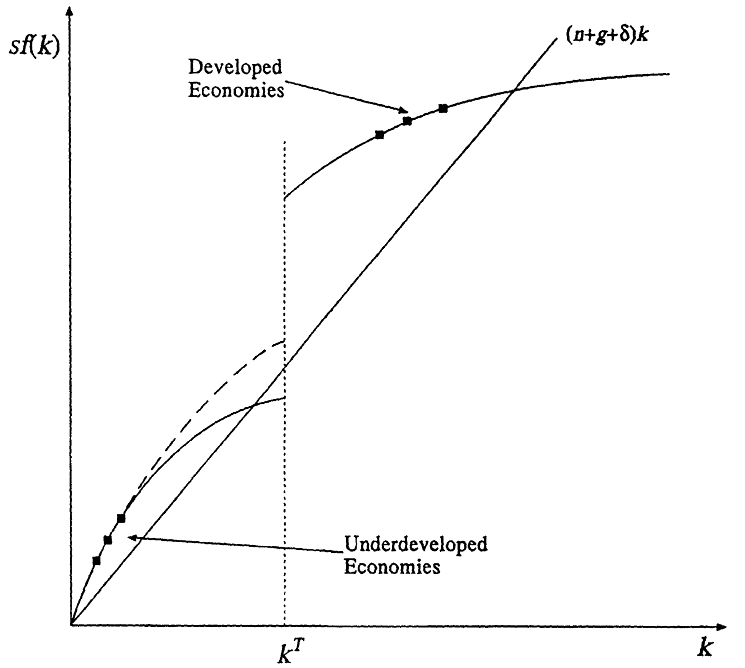

Figure 2. Multiple steady states versus stages of development

Cross-section data combining developed and undeveloped economies does not discriminate between the two candidate production functions.

Therefore, while our results show the compatibility of cross-country growth patterns with multiple steady states, they cannot be interpreted as a formal rejection of a single steady-state model. On the other hand, our evidence of multiple regimes is consistent with the notion that cross-country growth rates are affected by the way capital is converted on the margin into output regardless of whether one interprets our regimes as stages of development or as multiple steady states.

6.2. Residual Heterogeneity

One way to interpret our empirical procedure is to observe that while the standard linear growth regression rules out any heterogeneity in the growth process across countries, we allow for such heterogeneity across country groups. The regression tree procedure assumes all heterogeneity disappears once one sorts the economies into subgroups. While this assumption is justifiable in the context of certain multiple steady-state models, the assumption is nevertheless extreme as it rules out any country-specific differences. It is possible that each country obeys a regression of the form (2) with different coefficients. In this case, the regression tree procedure diminishes but does not eliminate heterogeneity in the cross-section regressions as it groups countries with similar laws of motion. As discussed in Pesaran and Smith (1994), this means that the within-group regression coefficients represent averages of the underlying individual coefficients for each country. As a result, our evidence of within-group convergence is compatible with some long-run differences between countries within a group.

6.3. Omitted Initial Conditions

While the use of initial income and literacy as conditioning variables produces country groupings which seem overall quite reasonable, there are some clear anomalies in the estimated regression tree. For example, Japan and the Republic of Korea are assigned to group 3 along with El Salvador whereas Trinidad and Tobago and Uruguay are assigned to group 4 along with the United States. These anomalies would seem most plausibly explained by the existence of additional initial conditions beyond those we study which are relevant for determining long-run growth patterns. One obvious candidate for such an omitted initial condition is 'social capital' (see Coleman, 1990, for a detailed discussion) which captures the role of cultural norms and values concerning interactions between individuals, which may range from attitudes towards work to respect for property rights, in economic growth. One obvious difficulty with a concept such as social capital is its lack of quantifiability, which indicates how econometric studies of the sort we have developed may be usefully augmented by country-specific studies.[20]

7. CONCLUSIONS

Taking as a starting place the work of Mankiw *et al.* (1992), we have re-examined the Summers–Heston data set to see whether the cross-country growth process exhibits multiple regimes in which subgroups of countries defined by initial conditions obey separate linear models. Our results are twofold. First, we reject the cross-country linear model specification which underlies most empirical work on growth. Second, we use regression tree methods to identify groups of countries which obey a common model. This analysis reveals substantial differences between the aggregate production functions of economies with different initial conditions. These features illustrate the compatibility of growth rate behaviour with a multiple steady-state perspective.

One important extension of our work is to see whether the multiple regimes we identify can be shown to arise from some of the production or demand complementarities which have been proposed as explanations for long-run divergence. The identification of these complementarities, in turn, will require a more careful understanding of within-country growth processes and thus seems likely to depend on the explicit analysis of a dynamic panel of countries. This line of research has been initiated in Quah (1992a,b); see also Pesaran and Smith (1994) for an analysis of many of the relevant econometric issues.

[20] Some evidence of the role of such factors may be found in Barro (1994), who finds correlations between different subjective measures of political freedom and growth.

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[b]	0·8	I	L
3	Benin	1116	5·0[b]	2·2	I	L
4	Botswana	959	na	8·6	I	na
5	Burkina Faso	529	2·0[b]	2·9	L	L
6	Burundi	755	14·0[b]	1·2	L	L
7	Cameroon	889	19·0[b]	5·7	I	L
8	Central African Republic	838	7·0[b]	1·5	I	L
9	Chad	908	6·0	−0·9	I	L
10	People's Republic of the Congo	1009	16·0[b]	6·2	I	L
11	Egypt	907	26·0	6·0	I	L
12	Ethiopia	533	15·0[a,b]	2·8	L	I
15	Ghana	1009	27·0[b]	1·0	I	L
17	Ivory Coast	1386	5·0[b]	5·1	I	L
18	Kenya	944	20·0[b]	4·8	I	L
20	Liberia	863	9·0[b]	3·3	I	L
21	Madagascar	1194	50·0[a,b]	1·4	I	I
22	Malawi	455	25·0[a,b]	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[b]	2·8	I	L
30	Rwanda	460	16·0[b]	4·5	L	L
31	Senegal	1392	6·0[b]	2·5	I	L
32	Sierra Leone	511	7·0[b]	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[b]	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[b]	5·6	I	L
40	Uganda	601	35·0[b]	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[b]	5·1	I	L
46	Bangladesh	846	22·0[b]	4·0	I	L
47	Burma	517	60·0[b]	4·5	L	I
48	Hong Kong	3085	70·0	8·9	I	H
49	India	978	28·0[b]	3·6	I	L
52	Israel	4802	84·0[b]	5·9	I	H
53	Japan	3493	98·0[b]	6·8	I	H
54	Jordan	2183	32·0[b]	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[b]	5·8	I	L
61	Philippines	1668	72·0	4·5	I	H

(continued)

DATA APPENDIX (*continued*)

Number	Country	$(Y/L)_{i,1960}$	$LR_{i,1960}$	Growth rate	Income class	Literacy class
63	Singapore	2 793	75·0[a,b]	9·2	I	H
64	Sri Lanka	1 794	75·0[b]	3·7	I	H
65	Syrian Arab Republic	2 382	30·0	6·7	I	L
67	Thailand	1 308	68·0	6·7	I	H
70	Austria	5 939	99·0	3·6	H	H
71	Belgium	6 789	99·0[a]	3·5	H	H
73	Denmark	8 551	99·0[a]	3·2	H	H
74	Finland	6 527	99·0[b]	3·7	H	H
75	France	7 215	99·0[a]	3·9	H	H
76	Federal Republic of Germany	7 695	99·0[a]	3·3	H	H
77	Greece	2 257	81·0	5·1	I	H
79	Ireland	4 411	97·0[b]	3·8	I	H
80	Italy	4 913	91·0[b]	3·8	I	H
83	Netherlands	7 689	99·0[a]	3·6	H	H
84	Norway	7 938	99·0[a,b]	4·3	H	H
85	Portugal	2 272	62·0	4·4	I	I
86	Spain	3 766	87·0	4·9	I	H
87	Sweden	7 802	99·0[a,b]	3·1	H	H
88	Switzerland	10 308	99·0[a]	2·5	H	H
89	Turkey	2 274	38·0	5·2	I	L
90	United Kingdom	7 634	99·0[a]	2·5	H	H
92	Canada	10 286	99·0[a]	4·2	H	H
93	Costa Rica	3 360	90·0[a]	4·7	I	H
94	Dominican Republic	1 939	65·0	5·1	I	I
95	El Salvador	2 042	49·0[b]	3·3	I	L
96	Guatemala	2 481	32·0	3·9	I	L
97	Haiti	1 096	15·0	1·8	I	L
98	Honduras	1 430	45·0[b]	4·0	I	L
99	Jamaica	2 729	82·0	2·1	I	H
100	Mexico	4 229	65·0	5·5	I	I
101	Nicaragua	3 195	57·0[a]	4·1	I	I
102	Panama	2 423	73·0	5·9	I	H
103	Trinidad and Tobago	9 253	93·0[b]	2·7	H	H
104	United States of America	12 362	98·0[b]	3·2	H	H
105	Argentina	4 852	91·0	2·1	H	H
106	Bolivia	1 618	39·0	3·3	I	L
107	Brazil	1 842	61·0	7·3	I	I
108	Chile	5 189	84·0	2·6	H	H
109	Columbia	2 672	63·0[b]	5·0	I	I
110	Ecuador	2 198	68·0[b]	5·7	I	H
112	Paraguay	1 951	75·0[b]	5·5	I	H
113	Peru	3 310	61·0	3·5	I	I
115	Uruguay	5 119	94·0[b]	0·9	H	H
116	Venezuela	10 367	63·0[b]	1·9	H	I
117	Australia	8 440	100·0[a]	3·8	H	H
119	Indonesia	879	39·0[b]	5·5	I	L
120	New Zealand	9 523	99·0[a]	2·7	H	H
121	Papua New Guinea	1 781	29·0	3·5	I	L

'Number' is the number given the country in the Summers and Heston (1988) data set.
na = not available.
[a] Literacy rate is for 1975 rather than 1960 as this is the next earliest available year.
[b] Literacy rate is for a year different, though no more than 2 years distant, from the specified year.

TECHNICAL APPENDIX: REGRESSION TREE ANALYSIS

This appendix describes the construction of a regression tree. The method can uncover general forms of nonlinearity in data; Brieman *et al.* (1984) show that the regression tree method is consistent in the sense that, under suitable regularity conditions, the estimated piecewise linear regression function converges to the best nonlinear predictor (in a mean squared error sense) of the dependent variable of interest.

Suppose that the optimal predictor of y_j given the vector $X_j = (x_{1,j}, \ldots, x_{r,j})$ is, in a mean squared error sense, the function $f(X_j)$. The estimation issue is the determination of $f(X)$ with little prior information on its form. One way of solving this problem is the following. Rewrite the support of each $x_{i,j}$ as the union of M intervals, $a_{i,0} \leq x_{i,j} < a_{i,1}, \ldots, a_{i,M-1} \leq x_{i,j} < a_{i,M}$. The support of X, S, can be expressed as the union of sets S_m, $m = 1 \ldots M^r$, each of which is a hyper-rectangle; for example S_1 can be defined as those elements of S such that $a_{i,0} \leq x_{i,j} < a_{i,1}$ for all i. The function $f(X)$ can then be approximated as a piecewise linear function of the form

$$f(X) \approx \sum_{m=1}^{M^r} \delta_m(X) X \beta_{S_m} \tag{A1}$$

where $\delta_m(X) = 1$ if $X \in S_m$, 0 otherwise and β_{S_m} is a constant vector. As the edges of each of the hyper-rectangles S_m are made small, this approximation can generally be made arbitrarily accurate.[21]

While the idea of estimating a piecewise linear approximation to $f(X)$ is appealing, one quickly runs into a curse of dimensionality problem if one were simply to search over the possible hyper-rectangle partitions of S so as to choose a particular piecewise linear approximation of $f(X)$. The problem is that the number of observations will quickly become very small relative to the number of hyper-rectangles in multiple dimensions, even for a small number of splits per individual variable. (See Härdle, 1990, for a discussion of this point.) Regression tree methods circumvent the curse of dimensionality by searching sequentially over the possible partitions of S.

The regression tree algorithm is as follows:

(1) For each of the variables x_i, $i = 1 \ldots r$, consider an initial split of the data into two subgroups according to the rule: assign observation j to $S_{(a,i)}$ if $x_{i,j} < a$, otherwise assign to $S_{(a^c,i)}$. Allowing a to range across the support of x_i traces out all possible binary splits of the data when x_i is the control. Repeating this for all i from 1 to r identifies all such binary splits. Let $\hat{\beta}_{(a,i)}$ denote the OLS estimate of the regression of y_j onto X_j using observations assigned to $S_{(a,i)}$; $\hat{\beta}_{(a^c,i)}$ is defined analogously. Some split variable x_i and some value a will minimize the sum of squared residuals (SSR):

$$\sum_{j \in S_{(a,i)}} (y_j - X_j \hat{\beta}_{(a,i)})^2 + \sum_{j \in S_{(a^c,i)}} (y_j - X_j \hat{\beta}_{(a^c,i)})^2 \tag{A2}$$

over all possible two-way splits. The x_i and a that minimize equation (A2) define the initial split of the data into two subgroups which we call S_1 and S_2; denote this set of splits as T_1.

(2) Repeat step 1 on each of the two subsets S_1 and S_2. The SSR minimizing split for observations in S_1 will define two new groups S_3 and S_4; S_5 and S_6 are constructed for observations in S_2. Notice that these new splits may occur on different variables, i.e. $j \neq k$.

[21] Notice that one can *a priori* set any of the elements of β_{S_m} equal to zero if one wants to use some variables for splitting without using them to predict within subgroups. We in fact do this for $LR_{i,1960}$.

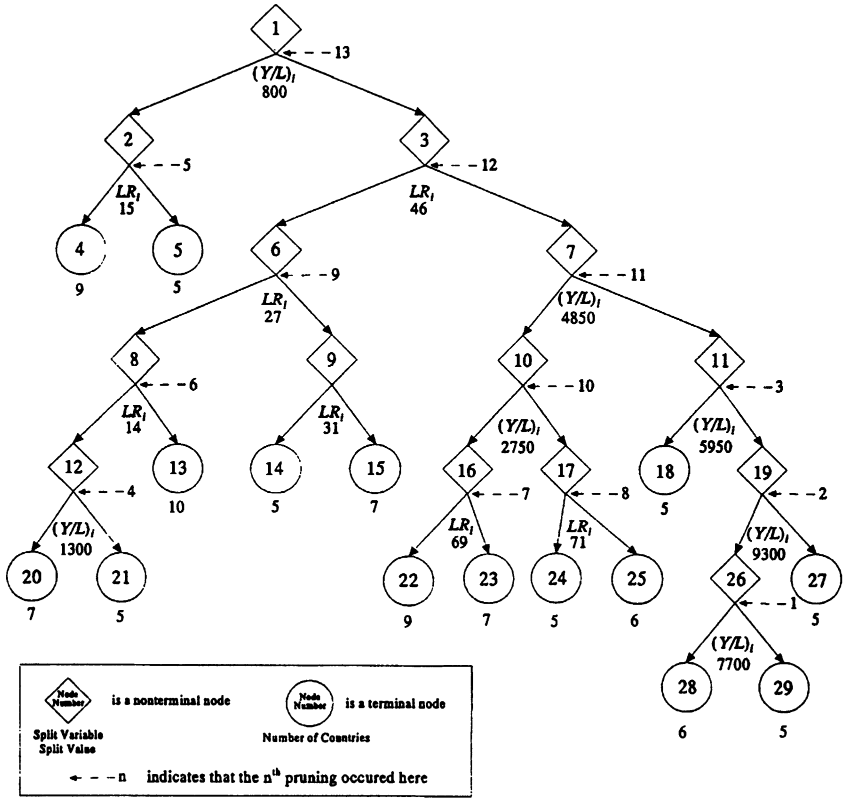

Figure 3. Unpruned regression tree

Denote this new set of splits $S_3 \ldots S_6$ as T_2. Repeat again for each of these new subsets and generate a new set of splits T_3. As before, the splits in T_3 define disjoint subgroups of data. Sequential splitting of each subset terminates either when there is no SSR reduction from splitting or when the number of observations in the cell is less than or equal to twice the number of regressors. Figure 3 illustrates this splitting for the Heston–Summers data. Let T_L denote the set of subsets S_k which are not further split; these terminal splits lie at the 'bottom' of the tree. The SSR for T_L equals

$$\sum_{S_m \in T_L} \sum_{j \in S_m} (y_j - X_j \hat{\beta}_{S_m})^2 \tag{A3}$$

(3) The piecewise linear model generated by step 2 is most likely overparameterized as the data splits were costless. We now 'prune' the tree which produced T_L by incorporating a cost to data splits. Let the cost of splitting equal $\alpha \cdot (\#(N) - 1)$, where $\#(N)$ is the number of terminal nodes in a tree. For each α, one determines which set of terminal nodes minimizes

$$\mathrm{SSR} + \alpha \cdot (\#(N) - 1) \tag{A4}$$

working backwards from T_L. First, remove any terminal splits in T_L whose elimination reduces the value of equation (A4), producing a new tree. Removal means the replacement of a pair of terminal nodes with a new terminal node which contains the set of observations whose split in the construction of T_L produced them. Next, remove all terminal splits of this new tree which, as before, are justified on the basis of minimizing equation (A4). Continue this sequential elimination of terminal nodes until no further removals will reduce equation (A4). This produces a tree with terminal nodes $T^*(\alpha)$. Constructing $T^*(\alpha)$ for all $0 \leqslant \alpha \leqslant \infty$ produces a series of trees and associated sets of terminal nodes starting with $T^*(0) = T_L$ that sequentially eliminates terminal splits so that $T^*(\infty)$ equals a single node which contains all observations. Figure 3 contains the tree used to produce our estimates in the text. Starting with an unpruned tree, increasing α from 0 first eliminates nodes 28 and 29, combining them to make node 26 terminal; further increasing α next combines nodes 27 and 28 to make node 19 terminal, etc.

(4) Calculate a cross-validated estimate of the SSR for each $T^*(\alpha)$. For each observation in the terminal splits of a given tree, form the predictor $\hat{\beta}_{-i,S_m}X_i$ for y_i where $\hat{\beta}_{-i,S_m}$ is the OLS estimate of β within subgroup S_m when the ith observation is omitted. Summing $(y_i - \hat{\beta}_{-i,S_m}X_i)^2$ over all observations produces the cross-validated SSR. The $T^*(\alpha)$ with the smallest cross-validated SSR produces the piecewise linear approximation which converges to the best nonlinear predictor.

The regression tree method resembles one which chooses among different splits using the Akaike information criterion. The key features of the tree approach are (1) splits are sequential, so that only a subset of all possible splits is examined, (2) cross-validation is used to assess model fit, (3) no penalty value is assigned *a priori*; rather, all possible penalties are considered.

ACKNOWLEDGEMENTS

This paper was originally circulated under the title 'Local versus global convergence across national economies'. We thank Robert Barro, James Hamilton, Blake LeBaron, Rodolfo Manuelli, Kenneth West, three anonymous referees and the editor Hashem Pesaran as well as seminar participants at Cambridge, Federal Reserve Board of Governors, London School of Economics, National Bureau of Economic Research, Oregon, Oxford, Rochester and Stanford for valuable comments. We especially thank Doug Steigerwald for initially suggesting the use of regression tree methods.

REFERENCES

Azariadis, C. and A. Drazen (1990), 'Threshold externalities in economic development', *Quarterly Journal of Economics*, **CV**, 501–526.

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

Barro, R. J. (1994), 'Democracy and growth', Working paper, National Bureau of Economic Research.

Barro, R. J. and H. C. Wolf (1989), 'Data appendix for economic growth in a cross-section of countries', Working Paper, National Bureau of Economic Research.

Baumol, W. J. (1986), 'Productivity growth, convergence and welfare: what the long run data show', *American Economic Review*, **76**, 1072–1085.

Bernard, A. B. and S. N. Durlauf (1993), 'Interpreting tests of the convergence hypothesis', Working paper, University of Wisconsin and forthcoming *Journal of Econometrics*.

Breiman, L., J. L. Friedman, R. A. Olshen and C. J. Stone (1984), *Classification and Regression Trees*, Wadsworth, Belmont, CA.

Coleman, J. S. (1990), *Foundations of Social Theory*, Harvard University Press, Cambridge, MA.

DeLong, J. B. (1988), 'Productivity growth, convergence and welfare: comment', *American Economic Review*, **78**, 1138–1154.

Dowrick, S. and N. Gemmell (1991), 'Industrialisation, catching up and economic growth: a comparative study across the world's economies', *Economic Journal*, **101**, 263–275.

Durlauf, S. N. (1993), 'Nonergodic economic growth', *Review of Economic Studies*, **60**, 349–366.
Härdle, W. (1990), *Applied Nonparametric Regression*, Cambridge University Press, Cambridge.
Lucas, R. E. (1988), 'On the mechanics of economic development', *Journal of Monetary Economics*, **22**, 3–42.
Mankiw, N. G., D. Romer and D. N. Weil (1992), 'A contribution to the empirics of economic growth', *Quarterly Journal of Economics*, **CVII**, 407–437.
Murphy, K., A. Shleifer and R. Vishny (1989), 'Industrialization and the big push', *Journal of Political Economy*, **97**, 1003–1026.
Pesaran, M. H. and R. Smith (1994), 'Estimating long-run relationships from dynamic heterogeneous panels', Working paper, University of Cambridge and forthcoming, *Journal of Econometrics*.
Quah, D. (1992a), 'International patterns of growth: I. Persistence in cross-country disparities', Working paper, London School of Economics.
Quah, D. (1992b), 'International patterns of growth: II. Persistence, path dependence and sustained take-off in growth transition', Working paper, London School of Economics.
Quah, D. (1993), 'Galton's fallacy and tests of the convergence hypothesis', *Scandinavian Journal of Economics*, **95**, 427–443.
Rauch, J. E. (1989), 'The question of per capita convergence in international consumption', Working paper, UC San Diego.
Romer, P. M. (1986), 'Increasing returns and long run growth', *Journal of Political Economy*, **94**, 1002–1037.
Summers, R. and A. Heston (1988), 'A new set of international comparisons of real product and prices: estimates for 130 countries', *Review of Income and Wealth*, **34**, 1–26.
White, H. (1980), 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica*, **48**, 817–838.