POLICY EVALUATION AND EMPIRICAL GROWTH RESEARCH

Steven N. Durlauf
University of Wisconsin at Madison

This paper explores the implications of the vast body of studies of cross-country growth determinants for the evaluation of alternative policies. Empirical growth studies have experienced a remarkable flowering in the last fifteen years, and innumerable insights have unquestionably been uncovered concerning similarities and differences in the growth experiences of various groups of countries. This empirical work was stimulated by—and, in turn, has been an essential complement to—the revival of growth theory initiated by the seminal papers of Lucas (1988) and Romer (1986). It constitutes one of the great successes of recent macroeconomic research.

In addition to identifying empirical regularities in growth, the empirical literature makes numerous claims concerning the impacts of alternative policies on the growth trajectories of different countries. This focus on policy implications is natural given the huge welfare implications of changes in a country's growth rate. A recent graduate-level textbook remarks

> "If large cross-sections of country experiences are interesting, it should mainly be because they ought to reveal the global impact of other growth determinants than the proximate factors of increases in productivity, factors about which we have other sources of evidence. Policy-oriented macroeconomists pay particular attention to the various components of government interventions" (Malinvaud, 1998, p. 781).

Durlauf and Quah (1999) survey the empirical growth literature and identify an enormous number of policy variables whose growth implications are analyzed in the new empirical literature. Among these

I thank Raimundo Soto for useful comments on a previous draft and Chih Ming Tan for outstanding research assistance. The John D. and Catherine T. MacArthur Foundation and Vilas Trust provided financial support.

Economic Growth: Sources, Trends, and Cycles, edited by Norman Loayza and Raimundo Soto, Santiago, Chile.

variables are government consumption (Barro, 1991), inflation (Barro, 1997), political instability (Alesina and others, 1996), civil liberties (Kormendi and Meguire 1985), financial repression (Easterly, 1993), tariffs (Lee, 1993), and trade openness (Harrison, 1995). And this list does not include variables such as human capital, for which the government's role is fundamental.

The argument of this paper, however, is that this empirical literature largely fails from the perspective of policy evaluation. Current econometric practice has yielded a body of evidence that is not policy relevant, in that a policymaker cannot readily translate the findings of the literature into implications for the evaluation of alternative policy trajectories. In making this argument, I focus on cross-country growth regressions of the type pioneered by Kormendi and Meguire (1985), Barro (1991), and Mankiw, Romer, and Weil (1992). While this style of empirical research does not exhaust the ways in which data have been brought to bear on growth questions, it does constitute the primary approach to empirical work in this literature. Furthermore, cross-country growth regressions have become a conventional mechanism through which policy recommendations are justified.

In what sense are cross-country growth regressions not policy relevant? I focus on two issues. First, when these regressions are used to make policy recommendations, the recommendations typically are based on the statistical significance of some regression coefficient. I argue that this way of using regressions does not have a natural decision-theoretic basis, because there is no simple relationship between statistical significance levels and policy evaluation. Second, growth regressions as conventionally constructed do not provide credible evidence of economic structure, so even if one is trying to use the regressions to solve a decision theory problem, it is unclear what information the regression actually contains.

It would be a gross caricature of the policymaking process to claim that there is a mechanical mapping from the statistical significance of certain regression parameters to specific policy decisions. The arguments in this essay apply more to the ways in which statistical evidence on growth are used for policy discussions among scholars. The two processes are, of course, linked. This essay describes more effective ways of translating statistical results into policy advice, so as to strengthen the contributions of academic discourse to policymaking.

My analysis is hardly the first critique of the empirical growth literature. Criticisms of the ways in which growth regressions are implemented and interpreted, as well as proposals for alternative approaches to analyzing growth data, may be found in Brock and Durlauf (2001), Durlauf (2000), Quah (1996, 1997), and Temple (2000), to name a few examples. The discussion here, of course, relies on my previous work, especially for technical justification and the development of the various arguments. In particular, much of the discussion represents an extension of Brock and Durlauf (2001). Relative to other critiques of growth empirics, my focus is on the specific failings of cross-county growth regressions in providing policy guidance. Nevertheless, many of the criticisms I make call into question not only whether evidence from growth regressions is actually informative for policy, but whether such regressions provide structural information on the sources of growth. I do not question the value of the empirical growth literature in terms of identifying stylized facts that theoretical models should address. Rather, the empirical growth literature fails when it moves in an insouciant fashion from stylized facts to causal claims.

I do not intend to be wholly nihilistic in this paper, and it is certainly not my belief that growth regressions have no place in the evaluation of policies. My goal is to highlight why claims based on such regressions should be modest. I also discuss some recent developments in statistics that I believe can enhance the utility of these regressions. In doing so, I strongly endorse two recent papers, Doppelhofer, Miller, and Sala-i-Martin (2000) and Fernandez, Ley, and Steel (2001b). My own previous work in Brock and Durlauf (2001) develops many of the arguments here at more length and constitutes a more formal statement of my views on how to conduct empirical growth analyses.

Section 1 discusses the basic question of growth regressions and policy evaluation, arguing that the appropriate link between the two processes is not reflected in conventional academic practice. What empirical work on growth should do is provide posterior densities for growth rates under alternative policy scenarios. I argue that the exercise of policy evaluation depends on these posterior densities, combined with an explicit statement of a policymaker's objectives. Section 2 then addresses why conventional approaches to growth econometrics do not provide credible estimates of the posterior densities needed for the type of policy evaluation exercise I advocate. The section argues that growth regressions suffer two basic problems: theory uncertainty and country heterogeneity. These problems have not been adequately addressed in the empirical growth studies. Section 3 describes a technique, Bayesian model averaging, that addresses the problems raised in section 2. Section 4 presents an empirical exercise to illustrate how Bayesian model averaging can influence the way one thinks about the effects of policies in light of empirical results. Finally, section 5 provides a summary and conclusions.

1. Regressions and Policy Analysis

This section describes the basics of growth regressions and suggests a general language for thinking about how regressions should influence policy evaluations. While growth regressions come in many forms, a canonical representation is

$$g_i = X_i\beta + Z_i\gamma + p_i\delta + \varepsilon_i\,, \tag{1}$$

where g_i is real per capita growth across some fixed time interval, X_i is a set of regressors suggested by the Solow growth model (population growth, technological change, physical and human capital, and savings rates, transformed in ways implied by the model), Z_i is a set of additional control variables suggested by new growth theories, p_i is the policy variable of interest, and ε_i is an error. The distinction between X_i and Z_i is important in econometric practice, because while X_i variables are essentially constant across empirical studies, there is no consensus on which Z_i variables should be included. Many growth studies use panel rather than cross-section data, but this difference has relatively little bearing on issues of interpretation and so is ignored here.[1]

What does it mean to use this type of regression to evaluate a policy? A policymaker presumably wishes to compare the effects of setting a policy variable at some fixed level, $\bar{p}$, with the effects of an alternative setting, $\bar{\bar{p}}$. Supposing that the policymaker has a payoff function,

$$V\left(y_i, R_i, p_i\right), \tag{2}$$

then the policy problem is essentially a comparison of the payoffs associated with the alternative policies. If the policymaker has a payoff function, V, the policy evaluation amounts to computing

$$EV\left(y_i, R_i, \bar{p} \middle| D\right) - EV\left(y_i, R_i, \bar{\bar{p}} \middle| D\right). \tag{3}$$

1. My argument is not intended to dismiss the utility of panels in studying growth across countries, but rather focuses on issues that apply in both cross-section and panel contexts. For example, while panels allow the elimination of fixed effects that correspond to constant differences in growth rates across countries, they do not provide any natural solution to the more general issue of parameter heterogeneity that I address below.

In this representation, E is an expected value operator, y_i represents per capita income in country i, R_i represents some set of characteristics of country i that affect the policymaker's assessments, and D denotes all data available to the policymaker. I have written the payoff function in terms of levels of per capita output, but if lagged per capita output is part of R_i, then this function can accommodate the case in which the growth rate is the relevant argument in the payoff function for the policymaker.

With regard to policy analysis, the key question is simple. How one can use regressions of the form in equation 1 to inform calculations of equation 3? This question is hardly an unusual one; indeed it is precisely this type of question that underlies the development of statistical decision theory, beginning with the seminal work of Abraham Wald (1950). From the Wald perspective, one evaluates a policy by calculating equation 3, using equation 1 to compute the conditional expectations that are a part of this calculation. Put differently, the relevance of equation 1 for policy analysis is that it allows for the computation of the distribution of growth rates under alternative choices of the policy variable, p_i. These distributions matter only in how they affect the expected payoff of the policymaker.

Surprisingly, this is not how policy implications are usually drawn from growth regressions. Instead, policy evaluations are drawn as an implication of hypothesis tests made on the coefficient associated with the policy variable of interest. In the context of equation 1, this amounts to using the statistical significance of δ in equation 1 to determine whether one can recommend a change in the magnitude of p_i to enhance growth in country i. A good example of this approach is the assessment of alternative policy variables in Barro and Sala-i-Martin (1995, chap. 12). In this survey of the empirical growth literature, the empirical evaluation of various policy variables in the growth process is virtually always related to statistical significance, typically assessed at the 5 percent level. (Significance at 10 percent but not 5 percent is apparently considered to be sufficiently weak evidence that a variable can be ignored.)

There is a vast statistical literature debating the use of statistical significance levels in evaluating statistical models; much of this debate revolves around frequentist versus Bayesian approaches to statistical analysis. My concern is somewhat different. The question is whether the statistical significance of a variable provides much insight into calculations of equation 3. As the form of equation 3 makes clear, the general answer is no. In order for there to be such a relationship, the payoff function would have to possess a functional form such that the implied policy recommendation would be, "implement the policy if the coefficient on the policy variable is statistically significant; otherwise do not implement the policy." Suppose that the question is whether to move from $\bar{p}$ to $\bar{\bar{p}}$. Assume that the ordinary least squares (OLS) estimate of δ, $\hat{\delta}$, can be interpreted as its expected value and that the OLS variance of the parameter estimate of δ is the variance of the parameter. (The conditions under which these assumptions hold are discussed below.) Finally, assume that the "statistical significance" rule is that one should only implement a policy change if the *t* statistic for the policy coefficient is greater than or equal to 2 and the sign of the policy change is the same as the coefficient estimate. Then, for the payoff function implicitly defined by

$$EV\left(y_i, R_i, \bar{p} \middle| D\right) - EV\left(y_i, R_i, \bar{\bar{p}} \middle| D\right) = E\left[\delta\left(\bar{p} - \bar{\bar{p}}\right)\right] - 2\operatorname{var}\left[\delta\left(\bar{p} - \bar{\bar{p}}\right)\right]^{1/2}, \tag{4}$$

one would only increase the policy variable from $\bar{p}$ to $\bar{\bar{p}}$ if the *t* statistic in the OLS regression is at least equal to 2 and the sign of the coefficient is positive. (The use of 2 versus some other value is immaterial.)

This is a very special case and embodies several unintuitive assumptions. First, the policymaker must only care about the component of growth affected by the control variable, rather than the effect of the control variable on growth per se. In other words, the policymaker considers the effect of the policy in isolation from all other determinants of growth. Second, the policymaker must only care about the mean and variance of the policy's effect on growth. While I do not wish to speculate on the objective functions employed in practice by policymakers, this function would seem inappropriate in many contexts. For example, political stability issues might render a policymaker more sensitive to negative growth rates than to positive growth rates, or there might be asymmetries in the effects of growth on poverty that should be accounted for in the social evaluation of changes in growth rates. Third, the mean and standard deviation of the growth effect in the payoff function must present a 2-to-1 tradeoff. This is where the significance level for the *t* statistic is implicitly embedded in the payoff function. The point, of course, is that there is no reason to expect any of these assumptions to hold in practice.

Is there a straightforward way to reduce the gap between the statistical decision theory approach to evaluating growth regressions and the use of statistical significance levels to evaluate policies? The message of the statistical decision theory literature is that no simple link exists. As demonstrated quite clearly in Chamberlain (2000), decision theory imposes powerful restrictions on how one analyzes data. Hence, one message for policymakers who must evaluate growth policies is that what they should ultimately care about is the posterior density of growth rates (or income levels) under alternative policy scenarios.

I therefore turn to a second question, namely, the interpretation of the posterior density of parameters in growth regressions. In interpreting the goal of an empirical exercise as the computation of a posterior density, I use Bayesian as opposed to frequentist language. This distinction is unimportant for the subsequent discussion, as the critiques I make of conventional growth regressions concern their interpretability, an issue that is equally salient under Bayesian and frequentist paradigms. Bayesian language is more appropriate in my discussion, however, because Bayesian approaches can be integrated much more naturally into decision-theoretic analyses than can frequentist approaches.

2. Interpreting Growth Regressions

The previous section illustrates why the conventional use of growth regressions for policy analysis has no logical justification. These regressions, nevertheless, do contain information about the growth process and its relationship to particular policy variables. In this section, I explore reasons why the regressions themselves are difficult to interpret in policy contexts.

The use of growth regressions to inform policy analyses is based on interpretations of these regressions as structural relationships. Put differently, these analyses presuppose that the observed correlations on which these regressions are computed can be interpreted as something more. Whether interpretations of the type found in this literature are justifiable is not entirely clear, for two main reasons: openendedness of the theories and parameter heterogeneity.

2.1 Openendedness and the Structure of Growth Theories

A first problem in specifying empirical growth models concerns the identification of the growth determinants to be included in a statistical model. This problem arises in any statistical analysis, but the danger is especially problematic in growth contexts. As originally argued in

Durlauf (2000), modern growth theories are fundamentally openended: one growth theory typically has no bearing on the empirical relevance of another. Modern growth economics has put forward an enormous range of alternative explanations for cross-country growth differences. Hence, one paper focuses on the effects of inequality on growth (Persson and Tabellini, 1994), another on the role of social capital (Knack and Keefer, 1997), another on geography, (Sala-i-Martin, 1997), and so on. What is critical in assessing individual empirical exercises is that these alternative explanations are both not mutually exclusive and often perfectly compatible. There is nothing in the logic of a theory linking social capital to growth that is inconsistent with a theory linking trade openness to growth, even though there may be interrelationships between the two theories.

How large is the set of growth theories that have been taken to data? Durlauf and Quah (1999) survey the empirical literature and note that as of 1998, at least as many ex ante plausible regressors have been used to proxy for growth theories as there are countries in the standard growth dataset. My own reading of the literature suggests that the number of potential variables has grown substantially since then.[2]

Openendedness has several critical implications for the interpretation of growth regressions. First, since the mutual compatibility of alternative growth theories in no way implies that they are uncorrelated (when their empirical analogues are compared across countries), the danger of omitted variable bias in a given cross-country regression is immense. Furthermore, the large number of growth variables means that one cannot simply run a regression with all theories, but rather must employ an empirical strategy for variable selection.

A number of studies make an effort to engage in variable selection in growth contexts and thereby deal with the dangers of misspecification. Levine and Renelt (1992) employ Edward Leamer's celebrated extreme bounds analysis to determine which variables can be robustly related to growth. This amounts to running a large set of regressions, each of which contains some subset of the potential regressors used for growth theories, and seeing how the sign of a given regressor changes according to what other regressors are included. When this sign is stable across alternative regressions, the regressor is considered to have a robust relationship. Sala-i-Martin (1997) employs a related procedure, but he interprets a regressor as robust if it is statistically significant in

2. I also believe that our 1999 survey seriously underestimates the number of growth theories that had been empirically employed up to that time.

95 percent of the regressions in which it is included. Each of these is an important paper. Neither can be said to provide a satisfactory resolution of the problem of variable selection, however, because neither approach has a fully satisfactory decision-theoretic foundation. Extreme bounds analysis implies that a policy variable is unimportant if it is not constant for a large number of regressions, although that sort of instability might well occur even if there is a relationship between the variable and growth. Furthermore, Levine and Renelt (1992) treat all regressions as equally informative, whereas standard metrics such as goodness of fit suggest they may not be. Similar criticisms may be made of Sala-i-Martin (1997). Brock and Durlauf (2001) provide some ways to think about decision-theoretic approaches to variable selection and discuss approaches that appear in the econometrics literature. For my purposes, however, the key point is that the proposed solutions to variable selection do not reflect the attention to decision-theoretic foundations needed to make the procedures wholly compelling. What I propose below incorporates the important insights of these techniques, but in a way that is more compatible with policy analysis.

Second, openendedness implies that it is extremely difficult to use instrumental variables in growth contexts. In regression 1, suppose one is worried that the policy variable, p_i, is endogenous. How can one construct an instrument for it? To be valid, the instrument must be predetermined with respect to p_i; this is the basis on which instruments are typically used in the growth literature. Validity also requires that a second condition be fulfilled: namely, that the instrument is uncorrelated with ε_i. And what is ε_i? This is the unobserved variable that captures all growth determinants that have not been modeled in the regression. Hence, to argue that the instrument is valid, one has to argue that it is uncorrelated with all theories not embodied by the regression, a condition that seems virtually impossible to satisfy.

Theory openendedness makes clear how prior information is an inevitable part of the interpretation of growth regressions and thus of their use in policy analysis. To interpret a particular regression of the form of equation 1 as revealing economic structure, it is necessary to believe something about the errors. As noted above, these errors embody growth determinants that have been neglected by the regression. The analyst must be able to interpret the parameter estimates despite the presence of these omitted variables. One's understanding of these omitted variables, however, reflects one's knowledge of the histories and societies of the various countries in the dataset. Such prior information often comes from qualitative and descriptive sources. The fact that these sources are not quantitative does not allow the analyst to ignore them.

2.2 Heterogeneity

A second issue in interpreting growth regressions concerns heterogeneity in countries. Put in its simplest form, the use of a regression such as equation 1 for policy analysis presupposes the belief that the growth process for different countries can be well approximated as a country-invariant relationship. Again, invariance of parameters across observations is an assumption that is certainly not unique to growth contexts, yet it seems particularly difficult to defend such an assumption in growth contexts. Consider the claim that a measure of the tariff level affects growth. Does one interpret the tariff coefficient in equation 1 as saying that the effects of a change in tariffs for the United States is the same as for Belgium and for Singapore? Does one believe that the growth implications of a unit change in human capital are the same for the United States as for countries in sub-Saharan Africa? Presumably not, but this is precisely what is asserted when one uses regressions such as equation 1 to uncover growth determinants.

Concerns about parameter heterogeneity are of more than theoretical interest. Studies such as Canova (1999), Desdoigts (1999), Durlauf and Johnson (1995), Durlauf, Kourtellos, and Minkin (2001), and Kourtellos (2000) all provide evidence of parameter heterogeneity. Taken as a whole, the evidence clearly suggests that the standard approach to assuming country-invariant parameters as the null modeling assumption in growth regressions is inconsistent with the data. This practice is still quite general, however, and most exceptions to this generalization amount to nothing more than ad hoc additions of cross products of growth variables with thresholds (for example, a variable that is zero for countries below some measure of income, 1 otherwise.) It is thus no exaggeration to say that parameter heterogeneity has yet to become a primary component of growth models.

From the perspective of evaluating growth policies, the implications of parameter heterogeneity are clear. Policy advice is not given in terms of average effects in the world. One would not say, “Since the average effect of tariffs on growth for all countries is negative, country *i* should lower tariffs.” Yet this is essentially what occurs when one neglects heterogeneity in the growth process and uses the coefficient estimates in equation 1 as the basis for advising a particular country.

Large deviations between the growth effect of a variable as estimated in a regression such as equation 1 and the effect for a particular country can easily occur when parameters are heterogeneous; there is no guarantee that the signs are even the same. Hence, neglected parameter heterogeneity can lead to bad policy advice.

3. Bayesian Model Averaging and Growth Regressions

This section describes a new approach to the analysis of growth regressions that has the potential for making inferences on growth regressions more credible. This technique is known as Bayesian model averaging (BMA). It has been developed by Adrian Raftery and a series of coauthors (Raftery, Madigan, and Hoeting, 1997; Hoeting, and others, 1999). Wasserman (2000) provides a very clear introduction to model averaging. Much of the motivation for the work can be traced to the issues of model uncertainty analyzed by Leamer (1978), a book whose importance to econometric practice has yet to be fully appreciated. Applications of BMA to growth regressions may be found in Brock and Durlauf (2001), Doppelhofer, Miller, and Sala-i-Martin (2000), and Fernandez, Ley, and Steel (2001b). While these papers differ in many details, each is motivated by similar concerns.

3.1 Basic Ideas

The basic idea of Bayesian model averaging is the following. Suppose that one is concerned with a parameter, in my case δ, that is an element of a model. Unlike conventional practice, however, suppose that one does not know the true model of which the parameter is an element. Instead, one has a set of models M, with a typical element M_m, which for the sake of exposition contains the true model. (If an element of the model set does not contain the parameter δ, it is interpreted as meaning $\delta = 0$).

Conventional econometric practice amounts to computing $\mu(\delta \mid D, M_m)$, that is, one evaluates the probability density of the parameter δ given the available data D and the assumption that the data are generated by a particular model M_m.[3] In contrast, Bayesian model averaging advocates computing the conditional probability of the parameter given only the data, that is, $\mu(\delta \mid D)$. This computation basi-

3. I sometimes refer to conditional probabilities as posterior probabilities, in recognition of their dependence on the available data.

cally eliminates the dependence of $\mu(\delta \mid D, M_m)$ on M_m by integrating out this additional conditioning variable. Since the number of models is discrete, this amounts to computing

$$\mu\left(\delta \middle| D\right) = \sum_m \mu\left(\delta \middle| D, M_m\right) \mu\left(M_m \middle| D\right). \tag{5}$$

Using Bayes rule, this expression may be rewritten as

$$\mu\left(\delta \middle| D\right) = \sum_m \mu\left(\delta \middle| D, M_m\right) \mu\left(D \middle| M_m\right) \mu\left(M_m\right), \tag{6}$$

which provides some insight into the difference between the BMA approach and conventional practice. Rather than condition on a single M_m in computing the posterior density, the BMA approach takes the posterior density $\mu(\delta \mid \mathrm{D}, M_m)$ for each model and computes a particular weighted average. The weights assigned to each model consist of two components: $\mu(M_m)$, which is the prior probability assigned to a given model, and $\mu(\mathrm{D} \mid M_m)$, which is the posterior probability of the data given a particular model. The latter term is nothing more than the likelihood function.

One can compute the posterior mean and variance of the parameter δ using these formulas. As originally shown in Leamer (1978), these are

$$E\left(\delta \middle| D\right) = \sum_m \mu\left(M_m \middle| D\right) E\left(\delta \middle| D, M_m\right) \text{ and} \tag{7}$$

$$\begin{aligned} \operatorname{var}\left(\delta \middle| D\right) &= E\left(\delta^2 \middle| D\right) - \left[E\left(\delta \middle| D\right)\right]^2 \\ &= \sum_m \mu\left(M_m \middle| D\right)\left\{\operatorname{var}\left(\delta \middle| M_m, D\right) + \left[E\left(\delta \middle| D, M_m\right)\right]^2\right\} - \left[E\left(\delta \middle| D\right)\right]^2 \\ &= \sum_m \mu\left(M_m \middle| D\right) \operatorname{var}\left(\delta \middle| M_m, D\right) \\ &\quad + \sum_m \mu\left(M_m \middle| D\right) \left[E\left(\delta \middle| D, M_m\right) - E\left(\delta \middle| D\right)\right]^2, \end{aligned} \tag{8}$$

respectively.

These formulas illustrate how model uncertainty affects a given parameter estimate. First, the posterior mean of the parameter is a weighted average of the posterior means across each model. Second, the posterior variance is the sum of two terms. The first term is a weighted average of the variances for each model. The second term reflects the variance across models of the expected value for δ; these differences reflect the fact that the models are themselves different. This second variance term captures how model uncertainty increases the variance associated with a parameter estimate relative to conventional calculations.[4]

3.2 Regression

From the perspective of a regression, model uncertainty is a function of what regressors to include. Both theory openendedness and country heterogeneity may be interpreted in this context. To see this, suppose that we have a set of R possible alternative determinants of growth. Theory openendedness means that the researcher does not have a basis for excluding one of the potential theories because another one matters. From the perspective of R, a regression that includes any subset of its elements constitutes a possible growth model. Hence, if there are K different regressors in R, then 2^{K-1} different possible models exist.

One can also interpret a range of possible forms of country heterogeneity in terms of variable inclusion in the same way I have interpreted theory uncertainty. Suppose that the countries in a growth cross section can be grouped into two distinct classes, such that countries within a class obey the same linear growth model. Let A_1 and A_2 denote the collections of country indices corresponding to these classes. The assumption of two classes means that there are two growth models for the data.

$$g_i = X_i\beta + Z_i\gamma + p_i\delta + \varepsilon_i \qquad \text{if } i \in A_1 \text{ and} \tag{9}$$

$$g_i = X_i\beta' + Z_i\gamma' + p_i\delta' + \varepsilon_i \qquad \text{if } i \in A_2\,. \tag{10}$$

Equation 10 can be rewritten, however, as

$$\begin{aligned} g_i = {} & X_i\beta + Z_i\gamma + p_i\delta + X_i\left(\beta' - \beta\right) \\ & + Z_i\left(\gamma' - \gamma\right) + p_i\left(\delta' - \delta\right) + \varepsilon_i \qquad \text{if } i \in A_2\,. \end{aligned} \tag{11}$$

4. See Draper (1995) for additional discussion.

Therefore, one can combine the data from both classes of countries into a single regression

$$g_i = X_i\beta + Z_i\gamma + p_i\delta + X_i\xi_{i,A_2}\left(\beta' - \beta\right) + Z_i\xi_{i,A_2}\left(\gamma' - \gamma\right) + p_i\xi_{i,A_2}\left(\delta' - \delta\right) + \varepsilon_i \qquad \text{if } i \in A_1 \cup A_2, \tag{12}$$

where ξ_{i,A_2} if $i \in A_2$, 0 otherwise. As this regression indicates, the presence of multiple classes of countries may be captured by introducing additional regressors, $X_i\xi_{i,A_2}$, $Z_i\xi_{i,A_2}$, and $p_i\xi_{i,A_2}$. It is straightforward to generalize this argument to multiple classes.

This approach is not completely general, in that it requires some prior judgments about what possible groups of countries will be considered. If one allows each country to have its own parameters, then one will not have enough degrees of freedom to estimate the model. I am currently unaware of any way to generalize BMA procedures to allow for endogenous determination of groups of countries with similar parameters. Such techniques could be developed, however. For example, regression tree methods of the type employed by Durlauf and Johnson (1995) to allow the data to reveal groups of similar countries could perhaps be incorporated into a BMA framework. This is a fruitful area for future research.

An alternative to this approach to parameter heterogeneity is to model the parameters of a growth regression as functions of country-specific characteristics. This would allow each country to be associated with a unique set of regression coefficients. See Durlauf, Kourtellos, and Minkin (2001) for an example of this approach. Both the approach here and the methods in Durlauf, Kourtellos, and Minkin (2001) can address issues of nonlinearities in the growth process; the differences between the approaches concerns how one approximates an unknown nonlinearity. The approach here has some advantage when one thinks that threshold effects are present, as suggested by models such as Azariadis and Drazen (1990); Galor (1996) provides additional discussion of issues related to nonlinearities in theory and empirical practice.

3.3 Implementation

To implement the BMA procedure, it is necessary to characterize $\mu(D \mid M_m)$ and $\mu(M_m)$. The former essentially requires the specification of two things: the prior distribution on the coefficients and the probability density for the residuals within a given model. I do this as follows. For a given regression, let S_i denote the regressor associated with country i. A growth regression will therefore have the form

$$g_i = S_i\zeta + \varepsilon_i \qquad i = 1,\ldots,I \ . \tag{13}$$

To compute the posterior distribution of ζ, I assume, first, that I have no informative prior information on the coefficients. In more standard language, I impose a noninformative prior on the coefficients, that is,

$$\mu(\zeta) \propto c \, . \tag{14}$$

Second, I assume that the errors are i.i.d. normal with a known variance. Under this assumption, one can show that the posterior density of the regression coefficients is

$$\mu(\zeta|D) \sim N\left[\hat{\zeta},(S'S)^{-1}\sigma_\varepsilon^2\right], \tag{15}$$

where $\hat{\zeta}$ is the OLS estimate of the parameters in equation 13.[5] Notice that $(S'S)^{-1}\sigma_\varepsilon^2$ is the OLS variance estimate for the parameters when the error variance is known. What is very useful about this formula is that it means that the parameters of the posterior density of ζ have OLS interpretations. The assumption that the error variance is known is not serious; the formula will still be valid asymptotically if σ_ε^2 is replaced with its OLS estimate.

The choice of $\mu(M_m)$ is more problematic, in that it corresponds to the prior information a researcher has about which model is true. At first glance, it might seem that if one does not have such information, one should assign equal prior weight to each element of M. This is not entirely satisfactory, however. Assigning equal probabilities to each model is equivalent to assuming that the prior probability of including a given regressor is 0.5 and is independent of the presence or absence of any other regressor. This is clearly untenable given the economics of growth. Presumably, the fact that inequality affects growth says something about whether political institutions affect growth. This observation is consistent with the problem of theory openendedness discussed earlier: mutual compatibility does not entail independence.

As yet, there has been no satisfactory proposal to deal with this problem. For the purposes of this paper, I employ the equal prior model

5. Compare the approach in Box and Tiao (1973, p. 115).

probability assumption. I do so, however, to allow for a simple interpretation of my results, not because it is intrinsically appealing. The development of better priors is an important outstanding research question. The approach I follow is in the spirit of the benchmark approach to choosing priors, which emphasizes the idea that priors should facilitate comparisons across studies. In this regard, Fernandez, Ley, and Steel (2001a) develop a number of priors for model averaging contexts that have desirable properties. Another approach is to use prior economic reasoning to structure priors. As argued in Brock and Durlauf (2001), the problem of interdependences in variables is analogous to violations of the assumption in discrete choice theory of the independence of irrelevant alternatives. This problem led to the development of models such as the nested logit; I conjecture that a similar tree structure may exist to organize growth regressors. Yet another possibility is the use of panel data as a training sample to form priors for the analysis of the rest of the sample. While this has yet to be done in the cases of variable selection, work such as that by Berger and Peracchi (2001, 2002) shows that this is a promising approach in other contexts.

4. Empirical Example

In this section, I describe a small BMA exercise in exploring the role of three different policy variables in growth. The exercise does not attempt to explore a large class of alternative theories, as would be ideal in analyses of this type and as is done in the important papers by Doppelhofer, Miller and Sala-i-Martin (2000) and Fernandez, Ley, and Steel (2001b). Rather, my intent is to illustrate how attention to model uncertainty can affect one's views of a regression.

My analysis focuses on a particular growth regression studied in an influential paper by Easterly and Levine (1997). I use their baseline regression, which is conducted on panel data constructed of ten-year moving averages for the decades 1960 to 1990.[6] Working with a baseline growth regression suggested by that paper, I consider various model averaging generalizations. Specifically, I focus on three policy variables that have received attention in the growth literature: a measure of the size of government deficits (SURP), a measure of human capital (SCHOOLING), and a measure of democracy (DEMOCRACY). One could plausibly argue that these are not control

6. Data definitions are provided in the appendix A below. See Easterly and Levine (1997) for additional discussion of the data.

variables with respect to any policymaker and should themselves be modeled as endogenous outcomes influenced by a policymaker, but this does not mitigate against the value of the exercise. (See the appendix B for estimation details.)

Six different regressions are run for each of these variables. First, I report an OLS growth regression that includes only the policy variable of interest. The regressor set referred to as ALL corresponds to the list of regressors in the first column of table 1. Second, I report a Bayesian model averaging exercise taken over all variables in the original model. Third, I report an ordinary least squares regression that allows for the coefficients on the Latin American countries to differ from the rest of the sample. The term LATINCA in the table refers to regressors that are set equal to zero for countries outside of Latin America. Fourth, I report the BMA analog to this regression. In the third and fourth cases, the regression coefficients I report are the ones that apply to the Latin American countries. Finally, I report OLS growth regressions and BMA analogues using only Latin American countries in columns 5 and 6. Columns 3 and 5, in principle, should be identical; differences here are second-order and reflect computational differences.

From the perspective of inferences about policy variables, the results of this exercise are mixed, in the sense that the conclusions one would draw from the BMA exercises are not systematically different from those that would be drawn from the OLS exercise. Nevertheless, there are some noteworthy differences. Perhaps the biggest difference concerns the SURP variable (see table 1). If one compares the estimates in columns 1 and 4 for this variable, one sees that the point estimate declines by about 30 percent and the standard error increases by about 30 percent when one engages in a BMA exercise that allows for both theory uncertainty and country heterogeneity. The exercise thus seems to undermine the evidence that this policy variable can be used to affect growth. Interestingly, the reason why the evidence is weakened is not simply that this analysis allows the Latin American countries to have different parameters than the rest of the world. As column 3 indicates, an application of BMA that only allows for theory uncertainty gives very similar results.

The impact of schooling on growth also exhibits some sensitivity to accounting for model uncertainty. Interestingly, this sensitivity is not uniform across alternative formulations. The posterior mean and standard deviation of the schooling parameter are quite similar when one compares columns 1 and 4, but they exhibit variations elsewhere. In particu-

Table 1. Government Deficit: Ordinary Least Squares versus Bayesian Model Averaging[a]

Explanatory variable	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Intercept term	—	—	—	—	–0.9084	–0.0333
	—	—	—	—	(0.3869)	(0.2305)
Dummy for sub-Saharan Africa	–0.0150	–0.0157	–0.0148	–0.0155	—	—
	(0.0044)	(0.0044)	(0.0045)	(0.0044)	—	—
Dummy for Latin America and the Caribbean	–0.0180	–0.0187	–0.8014	–0.0124	—	—
	(0.0037)	(0.0038)	(0.4425)	(0.0098)	—	—
Dummy for 1960s	–0.1428	0.0003	–0.1070	–0.0001	—	—
	(0.0831)	(0.0027)	(0.0863)	(0.0008)	—	—
Dummy for 1970s	–0.1388	0.0008	–0.1035	0.0008	0.0034	—
	(0.0829)	(0.0032)	(0.0866)	(0.0028)	(0.0060)	—
Dummy for 1980s	–0.1539	–0.0134	–0.1259	–0.0141	–0.0190	–0.0261
	(0.0828)	(0.0040)	(0.0867)	(0.0034)	(0.0068)	(0.0054)
Log of initial income	0.0559	0.0193	0.2446	0.0184	0.2446	0.0162
	(0.0215)	(0.0023)	(0.1108)	(0.0026)	(0.0987)	(0.0594)
Log of initial income squared	–0.0041	–0.0017	–0.0160	–0.0017	–0.0160	–0.0011
	(0.0014)	(0.0003)	(0.0071)	(0.0003)	(0.0063)	(0.0038)
Assassinations	–12.771	–0.6515	–22.1548	–0.8421	–22.1548	–4.8753
	(9.6661)	(3.4941)	(11.4928)	(3.9877)	(10.2465)	(9.1183)
Financial depth	0.0164	0.0114	0.0329	0.0113	0.0329	–0.0003
	(0.0059)	(0.0085)	(0.0260)	(0.0086)	(0.0231)	(0.0060)
Black market premium	–0.0204	–0.0226	–0.0133	–0.0221	–0.0133	–0.0090
	(0.0044)	(0.0045)	(0.0086)	(0.0047)	(0.0077)	(0.0099)
Ethnic diversity (ELF60)	–0.0189	–0.0208	–0.0019	–0.0201	–0.0019	–0.0016
	(0.0054)	(0.0055)	(0.0121)	(0.0064)	(0.0108)	(0.0057)
Fiscal surplus/GDP (SURP)	0.102	0.0775	0.1207	0.0792	0.1208	0.0458
	(0.0305)	(0.0417)	(0.0666)	(0.0401)	(0.0593)	(0.0573)

Source: Author's calculations, based on data from Easterly and Levine (1997).
a. The estimated regressions are as follows: (1) ordinary least squares (OLS) estimates for model ALL; (2) Bayesian model averaging (BMA) estimates for model ALL; (3) OLS estimates for model ALL + ALL*I(LATINCA), for which composite coefficient estimates and standard errors are reported and in which AFRICA, LATINCA, and DUM60 are dropped from the LATINCA-specific set of regressors; (4) BMA estimates for model ALL + ALL*I(LATINCA), for which composite coefficient estimates and standard errors are reported and in which AFRICA, LATINCA, and DUM60 are dropped from the LATINCA-specific set of regressors; (5) OLS on LATINCA subsample; and (6) BMA on LATINCA subsample. Standard errors are in parentheses.

lar, when BMA is applied to the Latin American countries in isolation (column 6), the posterior expected value of the schooling coefficient is less than half as large as the OLS estimate, with a much larger standard error as well. I am not sure how to interpret this finding.

Table 3 examines democracy and growth. This table fails to pick up any particularly interesting differences in the various democracy esti-

Table 2. Schooling: Ordinary Least Squares versus Bayesian Model Averaging[a]

Explanatory variable	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Intercept term	—	—	—	—	–0.7425	–0.0991
	—	—	—	—	(0.3837)	(0.3260)
Dummy for	–0.0144	–0.0160	–0.0146	–0.0159	—	—
sub-Saharan Africa	(0.0043)	(0.0043)	(0.0044)	(0.0043)	—	—
Dummy for Latin America	–0.0187	–0.0203	–0.6480	–0.0169	—	—
and the Caribbean	(0.0035)	(0.0036)	(0.4088)	(0.0136)	—	—
Dummy for 1960s	–0.1480	–0.00005	–0.0944	—	—	—
	(0.0813)	(0.00072)	(0.0863)	—	—	—
Dummy for 1970s	–0.1472	0.00004	–0.0892	0.0016	0.0052	0.0004
	(0.0814)	(0.00070)	(0.0865)	(0.0041)	(0.0059)	(0.0022)
Dummy for 1980s	–0.1630	–0.0152	–0.1146	–0.0155	–0.0201	–0.0259
	(0.0814)	(0.0027)	(0.0866)	(0.0035)	(0.0069)	(0.0059)
Log of initial income	0.0565	0.0190	0.1991	0.0184	0.1991	0.0316
	(0.0207)	(0.0022)	(0.1019)	(0.0029)	(0.0978)	(0.0837)
Log of initial income	–0.0044	–0.0019	0.0131	–0.0019	–0.0131	–0.0021
squared	(0.0013)	(0.0003)	(0.0064)	(0.0003)	(0.0062)	(0.0053)
Assassinations	–14.5187	–1.2223	–17.6708	–1.849	–17.6707	–5.8233
	(9.1125)	(4.6371)	(11.2981)	(5.7521)	(10.8444)	(9.9402)
Financial depth	0.0135	0.0072	0.0024	0.0073	0.0024	–0.0008
	(0.0055)	(0.0077)	(0.0196)	(0.0081)	(0.0188)	(0.0055)
Black market premium	–0.0230	–0.0242	–0.0191	–0.0240	–0.0192	–0.0129
	(0.0039)	(0.0039)	(0.0076)	(0.0039)	(0.0073)	(0.0096)
Ethnic diversity (ELF60)	–0.0160	–0.0194	–0.0079	–0.0191	–0.0079	–0.0010
	(0.0055)	(0.0055)	(0.0112)	(0.0057)	(0.0108)	(0.0045)
Log of SCHOOLING	0.0120	0.0107	0.0107	0.0114	0.0107	0.0018
	(0.0037)	(0.0047)	(0.0104)	(0.0048)	(0.0100)	(0.0053)

Source: Author's calculations, based on data from Easterly and Levine (1997).
a. The estimated regressions are as follows: (1) ordinary least squares (OLS) estimates for model ALL; (2) Bayesian model averaging (BMA) estimates for model ALL; (3) OLS estimates for model ALL + ALL*I(LATINCA), for which composite coefficient estimates and standard errors are reported and in which AFRICA, LATINCA, and DUM60 are dropped from the LATINCA-specific set of regressors; (4) BMA estimates for model ALL + ALL*I(LATINCA), for which composite coefficient estimates and standard errors are reported and in which AFRICA, LATINCA, and DUM60 are dropped from the LATINCA-specific set of regressors; (5) OLS on LATINCA subsample; and (6) BMA on LATINCA subsample. Standard errors are in parentheses.

mates, in that the posterior expectation in each case is quite small when compared with the variance. The consistency of this finding across the estimated alternatives, however, strengthens arguments that democracy levels do not seem to add much to empirical models of growth.[7]

7. Compare Barro (1996).

Table 3. Democracy: Ordinary Least Squares versus Bayesian Model Averaging[a]

Explanatory variable	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*
Intercept term	—	—	—	—	–0.7667	0.0083
	—	—	—	—	(0.4894)	(0.1529)
Dummy for sub-Saharan Africa	–0.0168	–0.0161	–0.0160	–0.0162	—	—
	(0.0047)	(0.0073)	(0.0049)	(0.0061)	—	—
Dummy for Latin America and the Caribbean	–0.0150	–0.0138	–0.6866	–0.0027	—	—
	(0.0045)	(0.0070)	(0.5295)	(0.0063)	—	—
Dummy for 1970s	–0.0949	0.0464	–0.0801	0.0467	—	—
	(0.0890)	(0.0155)	(0.0928)	(0.0146)	—	—
Dummy for 1980s	–0.1111	0.0293	–0.1052	0.0262	–0.0252	–0.0276
	(0.0890)	(0.0167)	(0.0930)	(0.0170)	(0.0062)	(0.0063)
Log of initial income	0.0423	–0.0001	0.2045	–0.0006	0.2045	0.0059
	(0.0231)	(0.0006)	(0.1304)	(0.0013)	(0.1224)	(0.0389)
Log of initial income squared	–0.0032	–0.0001	–0.0131	–0.0002	–0.0131	–0.0004
	(0.0015)	(0.0002)	(0.0082)	(0.0002)	(0.0077)	(0.0025)
Assassinations	–19.2393	–4.0676	–18.5783	–3.1335	–18.5783	–6.2317
	(10.3025)	(9.2535)	(12.4481)	(8.2420)	(11.6849)	(10.5951)
Financial depth	0.0172	0.0103	0.00780	0.0006	0.0079	–0.0019
	(0.0068)	(0.0115)	(0.0233)	(0.0187)	(0.0219)	(0.0087)
Black market premium	–0.0207	–0.0205	–0.0197	–0.0201	–0.0197	–0.0137
	(0.0044)	(0.0046)	(0.0087)	(0.0047)	(0.0081)	(0.0100)
Ethnic diversity (ELF60)	–0.0136	–0.0080	–0.0143	–0.0085	–0.0143	–0.0054
	(0.0059)	(0.0085)	(0.0148)	(0.0097)	(0.0140)	(0.0112)
DEMOCRACY	–0.0008	–0.000003	0.0005	—	0.0006	0.00003
	(0.0009)	(0.000190)	(0.0018)	—	(0.0016)	(0.00037)

Source: Author's calculations, based on data from Easterly and Levine (1997).
a. The estimated regressions are as follows: (1) ordinary least squares (OLS) estimates for model ALL; (2) Bayesian model averaging (BMA) estimates for model ALL; (3) OLS estimates for model ALL + ALL*I(LATINCA), for which composite coefficient estimates and standard errors are reported and in which AFRICA, LATINCA, and (LATINCA*DUM70) are dropped from LATINCA-specific set of regressors; (4) BMA estimates for model ALL + ALL*I(LATINCA), for which composite coefficient estimates and standard errors are reported and in which AFRICA, LATINCA, and (LATINCA*DUM70) are dropped from LATINCA-specific set of regressors; the democracy variable is included in set of regressors, but never picked up by BMA procedure. (5) OLS on LATINCA subsample; and (6) BMA on LATINCA subsample. Standard errors are in parentheses. No democracy data are available for the 1960s period, so DUM60 was dropped.

Finally, the tables generally reveal substantial differences for parameter estimates for Latin America versus the world as a whole. This strongly suggests that in using growth regressions to inform policy in Latin America, one must be very cautious in drawing generalizations from standard empirical exercises and applying them to Latin America. This finding is not unique; Brock and Durlauf (2001) draw similar conclusions with respect to countries in sub-Saharan Africa.

5. CONCLUSIONS

The objective of this paper was to argue that the conventional use of empirical studies of growth to inform policy suffers from a number of problems, which may be defined on two levels. First, the use of statistical significance levels to determine which policy instruments affect growth and which do not is an unsatisfactory basis for integrating empirical work and policy evaluation. Policy evaluation is better thought of as a comparison of posterior distributions of growth rates for a given country under alternative policy scenarios. This comparison can only be made relative to the payoff function of the policymaker. Statistical significance levels correspond to this comparison only for very special cases. Second, in assessing posterior distributions of growth rates, conventional growth regressions suffer from a number of limitations. These regressions typically do not allow for the fact that an empirical researcher does not know the true growth model. Model uncertainty, in this context, occurs because the modeler does not know what growth determinants must be included in a model or what forms of country-level heterogeneity need to be accounted for in the model. I follow Brock and Durlauf (2001), Doppelhofer, Miller, and Sala-i-Martin (2000), and Fernandez, Ley, and Steel (2001b) in advocating the use of Bayesian model averaging methods to allow for the explicit incorporation of model uncertainty in empirical work. A small empirical exercise illustrates how the use of growth regressions to draw policy implications for Latin American countries is affected by allowing for model uncertainty.

To repeat, this paper should not be regarded as advocating nihilism when it comes to econometric analyses of growth. Rather, it should be read as advocating caution. Within the discourse of academic economics, far too much emphasis is placed on zero-one assessments of whether a given theory is true. What is needed is a more nuanced approach to empirical work that gives adequate scope to the limits on inferences that can be made from observational data. Regressions have a role to play in policy evaluations, even for phenomena as important as growth. This role is distorted when a researcher ignores available historical and cultural information about a given country when conducting statistical work. Put differently, it is troubling how such deep qualitative studies as Greenfield (2001) and Landes (1998) have had little integration into the quantitative studies of growth that currently dominate the field.

The issues of the integration of econometrics with policy analysis and the appropriate incorporation of model uncertainty into empirical studies are by no means unique to the study of economic growth.

However, given the breadth of the phenomenon under study, as well as the complexities of the units whose behavior is to be evaluated (after all, we are dealing with the growth rates of entire economies), the study of growth at a country-wide level seems particularly susceptible to these problems. While there is no magic solution to the question of how to integrate different sources and types of information into a coherent policy exercise, such issues cannot be ignored. Ultimately, what is needed is a full recognition of the difficulties and limits facing any judgments that must be made in using data to inform growth policies.

APPENDIX A

The data cover 160 countries and include the following variables.

Code	*Description*
GYP	Growth rate of real per capita GDP. Source: *World Bank National Accounts* (various years).
AFRICA	Dummy variable for sub-Saharan African countries (according to World Bank definition). Source: *World Bank National Accounts* for AGO, BDI, BEN, BWA, CAF, CIV, CMR, COG, COM, CPV, DJI, ETH, GAB, GHA, GIN, GMB, GNB, GNQ, HVO, KEN, LBR, LSO, MDG, MLI, MOZ, MRT, MUS, MWI, NAM, NER, NGA, RWA, SDN, SEN, SLE, SOM, STP, SWZ, SYC, TCD, TGO, TZA, UGA, ZAF, ZAR, ZMB, ZWE (various years).
ASSASS	Number of assassinations per thousand population, decade average. Source: Banks (1994).
BLCK	Log of 1 plus black market premium, decade average. Source: World Bank (1991, with updates); *Pick's Currency Yearbook* (various years).
DUM60	Dummy variable for 1960s.
DUM70	Dummy variable for 1970s.
DUM80	Dummy variable for 1980s.
ELF60	Index of ethnolinguistic fractionalization, 1960. Measures the probability that two randomly selected people from a given country will not belong to the same ethnolinguistic group. Source: Easterly and Levine (1997); *Atlas Narodov Mira* (1964).
LATINCA	Dummy variable for Latin America and the Caribbean. Source: Easterly and Levine (1997).
LLY	Measure of financial depth, based on the ratio of the financial system's liquid liabilities to GDP, decade average. Liquid liabilities consist of currency held outside the banking system plus demand and interest bearing liabilities of banks and nonbank financial intermediaries. Source: King and Levine (1993).
LRGDP	Initial income, measured as the log of real per capita GDP at the start of each decade (1960, 1970, 1980). Source: Summers and Heston (1988).
LRGDPSQ	Log of initial real per capita GDP squared. Source: Summers and Heston (1988).
DEMOC	Measure of democracy (Gastil's political rights variable) Source: Gastil (1990, 1988).
LSCHOOL	Log of 1 plus average years of school attainment, quinquennial values (1960–1965, 1970–1975, and 1980–1985). Source: Barro and Lee (1993).
SURP	Ratio of central government fiscal surplus to GDP, both in local currency at current prices, decade average. Source: IMF's *International Financial Statistics* (various years, line 80) and *Government Finance Statistics* (various years, line L80).

APPENDIX B

All model averaging calculations were done using the program *bicreg*, which is an SPLUS program written by Adrian Raftery.[8] The key feature of the program is the way it deals with the large number of regressions involved in a BMA exercise. This program, following standard procedures in the model averaging literature, uses a search algorithm that explores only a subset of the model space; the design of the algorithm ensures that the search proceeds along directions such that it is likely to cover models that are relatively strongly supported by the data. I follow the procedure suggested by Madigan and Raftery (1995); see Raftery, Madigan, and Hoeting (1997) and Hoeting and others (1999) for additional discussion and a full description of the search algorithm. The latter paper provides a nice intuitive description of the ideas that underlie the algorithm:

> First, when the algorithm compares two nested models and decisively rejects the simpler model, then all submodels of the simpler model are rejected. The second idea, "Occam's window," concerns the interpretation of the ratio of posterior model probabilities $Pr(M_0/D)/Pr(M_1/D)$. Here M_0 is "smaller" than M_1.... If there is evidence for M_0 then M_1 is rejected, but rejecting M_0 requires strong evidence for the larger model M_1 (Hoeting and others, 1999, p. 385).

The algorithm I employ to implement the model averaging procedure uses an approximation, following Raftery (1995), based on the idea that for a large enough number of observations, the posterior coefficient distribution will be close to the maximum likelihood estimator, such that one can use the maximum likelihood estimates to avoid the need to specify a particular prior. Raftery (1995) and Tierney and Kadane (1986) contain technical details. While some evidence exists that this approximation works well in practice, more research is needed on the specification of priors for model averaging. Fernandez, Ley, and Steel (2001a) make an important contribution in this respect.

8. Available at www.research.att.com/~volinsky/bma.html.

REFERENCES

Alesina, A., and others. 1996. "Political Stability and Economic Growth." *Journal of Economic Growth* 1(2): 189–211.

Atlas Narodov Mira. 1964. Moscow: Miklukho-Maklai Ethnological Institute at the Department of Geodesy and Cartography of the State Geological Committee of the Soviet Union.

Azariadis, C., and A. Drazen. 1990. "Threshold Externalities in Economic Development." *Quarterly Journal of Economics* 105(2): 501–26.

Banks, A. 1994. "Cross-National Time Series Data Archive." State University of New York at Binghamton, Center for Social Analysis.

Barro, R. 1991. "Economic Growth in a Cross-Section of Countries." *Quarterly Journal of Economics* 106(2): 407–43.

————. 1996. "Democracy and Growth." *Journal of Economic Growth* 1(1): 1–27.

————. 1997. *Determinants of Economic Growth*. MIT Press.

Barro, R., and J.-W. Lee. 1993. "International Comparisons of Educational Attainment." *Journal of Monetary Economics* 32(3): 363–94.

Barro, R., and X. Sala-i-Martin. 1995. *Economic Growth*. New York: W. W. Norton.

Berger, J., and L. Perrachi. 2001. "Objective Bayesian Methods for Model Selection: Introduction and Comparison." In *Model Selection*, edited by P. Lahiri. *Lecture Notes Monograph Series*, vol. 38. Beachwood, Ohio: Institute for Mathematical Statistics.

————. 2002. "Training Samples in Objective Bayesian Model Selection." Duke University, Department of Statistics. Mimeographed.

Box, G., and G. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. New York: John Wiley and Sons.

Brock, W., and S. Durlauf. 2001. "Growth Economics and Reality." *World Bank Economic Review* 15(2): 229–72.

Canova, F. 1999. "Testing for Convergence Clubs in Cross-Country Growth Data: A Predictive Density Approach." University of Pompeu at Fabra (Spain), Department of Economics.

Chamberlain, G. 2000. "Econometrics and Decision Theory." *Journal of Econometrics* 95(2): 255–83.

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

Doppelhofer, G., R. Miller, and X. Sala-i-Martin. 2000. "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach." Working paper 7750. Cambridge, Mass.: National Bureau of Economic Research.

Draper, D. 1995. “Assessment and Propagation of Model Uncertainty.” *Journal of the Royal Statistical Society* 57 (series B): 45–70.
Durlauf, S. 2000. “Econometric Analysis and the Study of Economic Growth: A Skeptical Perspective.” In *Macroeconomics and the Real World*, edited by R. Backhouse and A. Salanti. Oxford University Press.
Durlauf, S., and P. Johnson. 1995. “Multiple Regimes and Cross-Country Growth Behavior.” *Journal of Applied Econometrics* 10,(4): 363–84.
Durlauf, S., A. Kourtellos, and A. Minkin. 2001. “The Local Solow Growth Model.” *European Economic Review* 45 (May): 928–40.
Durlauf, S., and D. Quah. 1999. “The New Empirics of Economic Growth.” In *Handbook of Macroeconomics*, edited by J. Taylor and M. Woodford. Amsterdam: North-Holland.
Easterly, W. 1993. “How Much Do Distortions Affect Growth?” *Journal of Monetary Economics* 32(2): 187–212.
Easterly, W., and R. Levine. 1997. “Africa’s Growth Tragedy: Policies and Ethnic Divisions.” *Quarterly Journal of Economics* 112(4): 1203–50.
Fernandez, C., E. Ley, and M. Steel. 2001a. “Benchmark Priors for Bayesian Model Averaging.” *Journal of Econometrics* 100(2) 381–427.
———. 2001b. “Model Uncertainty in Cross-Country Growth Regressions.” *Journal of Applied Econometrics* 16(5): 563–76.
Galor, O. 1996. “Convergence? Inferences from Theoretical Models.” *Economic Journal* 106(437): 1056–69.
Gastil, R. 1988. *Freedom in the World: Political Rights and Civil Liberties, 1987–1988*. Washington: Freedom House.
———. 1990. *Freedom in the World: Political Rights and Civil Liberties, 1989–1990*. Washington: Freedom House.
Greenfield, L. 2001. *The Spirit of Capitalism*. Harvard University Press.
Harrison, A. 1995. “Openness and Growth: A Time-Series, Cross-Section Analysis for Developing Countries.” Working paper 5221. Cambridge, Mass.: National Bureau of Economic Research.
Hoeting, J., and others. 1999. “Bayesian Model Averaging: A Tutorial.” *Statistical Science* 14(4): 382–401.
IMF (International Monetary Fund). Various years. *Government Finance Statistics*. Washington.
———. Various years. *International Financial Statistics*. Washington.
King, R., and R. Levine. 1993. “Finance, Entrepreneurship, and Growth: Theory and Evidence.” *Journal of Monetary Economics* 32(3): 513–42.
Knack, S., and P. Keefer 1997. “Does Social Capital Have an Economic Payoff? A Cross-Country Investigation.” *Quarterly Journal of Economics* 112(4): 1251–88.

Kormendi, R., and P. Meguire. 1985. "Macroeconomic Determinants of Growth: Cross-Country Evidence." *Journal of Monetary Economics* 16(2): 141–63.

Kourtellos, A. 2000. "A Projection Pursuit to Cross-Country Growth Data." University of Wisconsin at Madison. Mimeographed.

Landes, D. 1998. *The Wealth and Poverty of Nations.* New York: W. W. Norton.

Leamer, E. 1978. *Specification Searches.* New York: John Wiley and Sons.

Lee, J.-W. 1993. "International Trade, Distortions, and Long-Run Economic Growth." *IMF Staff Papers* 40 (June): 299–328.

Levine, R., and D. Renelt. 1992. "A Sensitivity Analysis of Cross-Country Growth Regressions." *American Economic Review* 82(4): 942–63.

Lucas, R. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22(1): 3–42.

Madigan, D., and A. Raftery. 1994. "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window." *Journal of the American Statistical Association* 89(428): 1535–46.

Malinvaud, E. 1998. *Macroeconomic Theory, vol. B: Economic Growth and Short-Run Equilibrium.* Amsterdam: North-Holland.

Mankiw, N. G., D. Romer, and D. Weil. 1992. "A Contribution to the Empirics of Economic Growth." *Quarterly Journal of Economics* 107(2): 407–37.

Persson, T., and G. Tabellini. 1994. "Is Inequality Harmful for Growth?" *American Economic Review* 84(3): 600–21.

Pick Publishing Corporation. Various years. *Pick's Currency Yearbook.* New York.

Quah, D. 1996. "Empirics for Growth and Convergence." *European Economic Review* 40(6): 1353–75.

———. 1997. "Empirics for Growth and Distribution: Polarization, Stratification, and Convergence Clubs." *Journal of Economic Growth* 2(1): 27–59.

Raftery, A. 1995. "Bayesian Model Selection in Social Research (with discussion)." In *Sociological Methodology* edited by P. V. Marsden, 111–95. Cambridge, Mass.: Blackwell.

Raftery, A., D. Madigan, and J. Hoeting. 1997. "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association* 92(437): 179–91.

Romer, P. 1986. "Increasing Returns and Long Run Growth." *Journal of Political Economy* 94(5): 1002–37.

Sala-i-Martin, X. 1997. "I Just Ran Two Million Regressions." *American Economic Review* 87(2): 178–83.

Summers, R., and A. Heston. 1988. "A New Set of International Comparisons of Real Product and Price Levels Estimates for 130 Countries, 1950–1985." *Review of Income and Wealth* 34(1): 1–25.

Temple, J. 2000. "Growth Regressions and What the Textbooks Don't Tell You." *Bulletin of Economic Research* 52(3): 181–205.

Tierney, L., and J. Kadane. 1986. "Accurate Approximations for Posterior Moments and Marginal Densities." *Journal of the American Statistical Association* 81(393): 82–86.

Wald, A. 1950. *Statistical Decision Functions.* New York: John Wiley and Sons.

Wasserman, L. 2000. "Bayesian Model Selection and Bayesian Model Averaging." *Journal of Mathematical Psychology* 44(1): 97–112.

World Bank. 1991. *World Development Report 1991: The Challenge of Development.* Oxford University Press.

———. Various years. *World Bank National Accounts.* Washington.