What have we learned from a decade of empirical research on growth?

Growth Empirics and Reality

William A. Brock and Steven N. Durlauf

This article questions current empirical practice in the study of growth. It argues that much of the modern empirical growth literature is based on assumptions about regressors, residuals, and parameters that are implausible from the perspective of both economic theory and the historical experiences of the countries under study. Many of these problems, it argues, are forms of violations of an exchangeability assumption that implicitly underlies standard growth exercises. The article shows that these implausible assumptions can be relaxed by allowing for uncertainty in model specification. Model uncertainty consists of two types: theory uncertainty, which relates to which growth determinants should be included in a model; and heterogeneity uncertainty, which relates to which observations in a data set constitute draw from the same statistical model. The article proposes ways to account for both theory and heterogeneity uncertainty. Finally, using an explicit decision-theoretic framework, the authors describe how one can engage in policy-relevant empirical analysis.

There are more things in heaven and earth, Horatio,
Than are dreamt of in your philosophy.   —William Shakespeare
                                                     Hamlet, act 1, scene 5

The objective of this article is ambitious—to outline a perspective on empirical growth research that will both address some of the major criticisms to which this research has been subjected and facilitate policy-relevant empirics. It is no exaggeration to say that the endogenous growth models pioneered in Romer (1986, 1990) and Lucas (1988) have produced a sea change in the sorts of questions around which macroeconomic research is focused. In empirical macroeconomics, efforts to explain cross-country differences in growth behavior since World War II have become a predominant area of research. The implications of this work for policymakers are immense. For example, strong links exist between national growth performance and international poverty and inequality. Differ-

ences in per capita income across countries are substantially larger than those within countries; Schultz (1998) concludes that two-thirds of (conventionally measured) inequality across individuals internationally is due to intercountry differences, so that efforts to reduce international inequality naturally focus on cross-country growth differences. In turn, the academic community has used this new empirical work as the basis for strong policy recommendations. A good example is Barro (1996). Based on a linear cross-country growth regression of the type so standard in this literature, Barro (1996, p. 24) concludes that

> The analysis has implications for the desirability of exporting democratic institutions from the advanced western economies to developing nations. The first lesson is that more democracy is not the key to economic growth. . . . The more general conclusion is that advanced western countries would contribute more to the welfare of poor nations by exporting their economic systems, notably property rights and free markets, rather than their political systems.

Yet there is widespread dissatisfaction with conventional empirical methods of growth analysis. Many critiques of growth econometrics have appeared in recent years. Typical examples include Pack (1994, pp. 68–69) who described a litany of problems with cross-country growth regressions:

> Once both random shocks and macroeconomic policy variables are recognized as important, it is no longer clear how to interpret many of the explanations of cross-country growth. . . . Many of the right hand side variables are endogenous. . . . The production function interpretation is further muddled by the assumption that all countries are on the same international production frontier . . . regression equations that attempt to sort out the sources of growth also generally ignore interaction effects. . . . The recent spate of cross-country growth regressions also obscures some of the lessons that have been learned from the analysis of policy in individual countries.

Another is Schultz (1999, p. 71): "Macroeconomic studies of growth often seek to explain differences in economic growth rates across countries in terms of levels and changes in education and health human capital, among other variables. However, these estimates are plagued by measurement error and specification problems." In fact, it seems no exaggeration to say that the growth literature in economics is notable for the large gaps that persist between theory and empirics. A recent (and critical) survey of the empirical literature, Durlauf and Quah (1999, p. 295), concludes that

> the new empirical growth literature remains in its infancy. While the literature has shown that the Solow model has substantial statistical power in explaining cross-country growth variation, sufficiently many problems exist with this work that the causal significance of the model is not clear. Fur-

ther, the new stylized facts of growth, as embodied in nonlinearities and distributional dynamics have yet to be integrated into full structural econometric analysis.

Our purposes in this article are threefold. First, we attempt to identify some general methodological problems that we believe explain the widespread mistrust of growth regressions. Although the factors we identify are not exhaustive, they do represent many of the most serious criticisms of conventional growth econometrics of which we are aware. These problems are important enough to at best seriously qualify and at worst invalidate many of the standard claims made in the new growth literature concerning the identification of economic structure. In particular, we argue that causal inferences as conventionally drawn in the empirical growth literature require certain statistical assumptions that may easily be argued to be implausible. This assertion holds from the perspective of both economic theory and the historical experiences of the countries under study. We further argue that a major source of skepticism about the empirical growth literature, and one that incorporates many of the usual criticisms, is the failure of certain statistical conditions representing forms of a property known as exchangeability to hold in conventional empirical growth exercises.

Second, we argue that the exchangeability failures underlying many criticisms of growth models may be constructively dealt with through explicit attention to model uncertainty in the formulation of growth regressions; see Temple (2000) for a complementary analysis. What we mean is the following: In estimating a particular regression, the inferences are made conditional both on the data and on the specification of the regression. The exchangeability objection to a regression amounts to questioning whether the specification of the regression is correct. The assumption that a particular specification is correct can be relaxed by treating model specification as an additional unknown feature of the data, that is, by explicitly incorporating model uncertainty in the statistical analysis. In taking this approach, we follow some important recent developments in the empirical growth literature—Fernandez, Ley, and Steel (1999) and Doppelhofer, Miller, and Sala-i-Martin (2000)—in endorsing the use of Bayesian methods to address explicitly the model uncertainty that we believe underlies the mistrust of conventional growth regressions. This analysis does not address all the criticisms we describe in the first part of the article. In particular, we argue that questions of causality versus correlation, which are of first-order importance in interpreting growth regressions, may only be addressed using substantive information that originates outside the models under analysis. Nevertheless, accounting for exchangeability can strengthen the confidence that may be attached to causal interpretations of regression exercises.

Third, we argue that the appropriate use of empirical growth analyses for policy analysis requires an explicit decision-theoretic formulation. Current empirical practice in growth is therefore not "policy-relevant" in the sense that the

policy inferences of a given data analysis are decoupled from the analysis itself. For example, one often sees a statistically insignificant coefficient used as evidence that some policy is not important for growth or, conversely, the assertion that statistical significance establishes the importance of some policy. We argue that these types of claims are not appropriate. Ideally, empirical growth exercises should employ cross-country growth data to compute predictive distributions for the consequences of policy outcomes, distributions that can then be combined with a policymaker's welfare function to assess alternative policy scenarios. A decision-theoretic approach to evaluating growth regressions can provide a better measure of the level of the evidence inherent in the available data, especially for the construction of policy-relevant predictive structures through empirical growth analyses.

The title of this article intentionally echoes the classic Sims (1980) critique of macroeconometric models. The growth literature does not suffer from the exact type of "incredible" assumptions (Sims 1980) that were required to identify economic structure through 1960s-style simultaneous equation models and whose interpretation Sims was attacking. Yet this literature does rely on assumptions that may be argued to be equally dubious and whose implausibility renders the inferences typically claimed by empirical workers to be equally suspect.[1] As will be clear from our discussion, this article only begins to scratch the surface of a policy-relevant growth econometrics. Our hope is that the ideas herein will facilitate new directions in growth research.

At the same time, our purpose is not to argue that statistical analyses of cross-country growth data are incapable of providing insights. Regression and other forms of statistical analysis have several critical roles in the study of growth. One role is the identification of interesting data patterns, patterns that can both stimulate economic theory and suggest directions along which to engage in country-specific studies. Quah's work (1996a, 1996b, 1997) is exemplary in this regard. However, we focus explicitly on the role of empirical work in formulating policy recommendations. In particular, a second goal of this article is to explore how one can, by casting empirical analysis in an explicitly decision-theoretic framework, develop firmer insights into the growth process.

Throughout we will take an eclectic stance on how one should go about data analysis. Many of our ideas are derived from the Bayesian statistics literature. Yet the basic arguments we make are relevant to frequentist analyses. Our view of data analysis is essentially pragmatic. Data analyses of the sort that are conventional in economics should be thought of as evidence-gathering exercises aimed at facilitating the evaluation of hypotheses and the development of policy-relevant predictions for future trajectories of variables of interest. For example, one starts with a proposition such as "the level of democracy in a country causally influences the level of economic growth." Once this statement is mathemati-

---

1. A number of the issues we raise echo, at least in spirit, Freedman (1991, 1997), who has made serious criticisms of the use of regressions to uncover causal structure in the social sciences.

cally instantiated (which means that *ceteris paribus* conditions are formalized, a more or less convincing theoretical model or set of models of causal influence is formulated in a form suitable for econometric implementation, etc.), the purpose of an empirical exercise is to see whether the statement is more or less plausible once the analysis has been conducted. The success or failure of an empirical exercise rests on whether one's prior views of the proposition have been altered by the analysis and on whether the level of uncertainty around a conclusion is low enough for the conclusion to be of policy relevance. Our position is that one should evaluate statistical procedures on the basis of whether they successfully answer the questions for which they are employed; we are unconcerned, at least in this article, with abstract issues that distinguish frequentist and Bayesian approaches, for example.

Many of our criticisms of the empirical growth literature apply in principle to other empirical contexts. They take on particular force in the growth context because of the complexity of the objects under study, the poor data available for empirical growth work, and the qualitative nature of the theories that drive the new growth literature.

I. A Baseline Regression

The bulk of modern empirical work on growth has focused on cross-country growth regressions of the type pioneered by Barro (1991) and Mankiw, Romer, and Weil (1992). Although recent work has extended growth analysis to consider panels (Evans 1998; Islam 1995; Lee, Pesaran, and Smith 1997), the arguments we make relating to conventional empirical growth practice as well as our proposed alternative approach are generally relevant to that context too, so long as cross-section variation is needed for parameter identification. Hence we focus on cross-sections.

A generic form for various cross-country growth regressions is

$$(1) \qquad g_i = X_i\gamma + Z_i\pi + \varepsilon_i$$

where g_i is real per capita growth in economy i over a given period (typically measured as the change in per capita income between the beginning and end of the sample divided by the number of years that have elapsed). We have divided the regressors into two types. X_i represents variables whose presence is suggested by the Solow growth model: a constant, initial income and a set of country-specific savings and population growth rate controls. The Solow model is often treated as a baseline from which to build up more elaborate growth models, hence these variables tend to be common across studies. Z_i, in contrast, consists of variables chosen to capture additional causal growth determinants that a researcher believes are important and so generally differs across analyses.[2]

---

2. See Galor (1996) for a discussion of the implications of different growth theories for convergence and Bernard and Durlauf (1996) for an analysis of the economic and statistical meanings of convergence.

Though this regression is typically applied to national aggregates, it can in principle be applied to regions or sectors once g_i is reinterpreted as a vector of growth rates within a country. This is particularly important for policy analysis when a given policy may affect different regions or population groups differently. Our conjecture is that such decompositions are important when evaluating growth policies with significant distributional consequences.

In our discussion we assume that the motivation for the estimation of a regression of the type given in equation 1 is policy driven. Specifically, we assume that a policymaker is interested in using this equation to advise some country i on whether it should change some policy instrument z. If the policymaker's objective function depends on the growth rate in country i, he will presumably need to understand the country's overall growth process and hence to make inferences about a number of aspects of equation 1 in addition to π_z, the coefficient on the policy instrument. We return to this issue in section VI.

II. ECONOMETRIC ISSUES

In this section we discuss three problems with the use of the baseline equation 1 in policymaking or other exercises in which one wishes to give a structural interpretation to this regression. These problems all, at one level, occur because of violations of the assumptions necessary to estimate equation 1 using ordinary least squares (OLS) and interpret the estimated equation as the structural model of growth dynamics implied by the augmented Solow model. Each of these criticisms ultimately reduces to questioning whether growth regressions as conventionally analyzed can provide the causal inferences that motivate such analyses. As discussed in the introduction, growth regressions have been subjected to a wide range of criticisms from many authors. We do not claim that any of the criticisms are necessarily original to us; instead, we believe our contribution in this section lies in the way we organize and unify these criticisms.

Open-Endedness of Theories

A fundamental problem with growth regressions is determining what variables to include in the analysis. This problem occurs because growth theories are open-ended. By open-endedness, we refer to the idea that the validity of one causal theory of growth does not imply the falsity of another. So, for example, a causal relationship between inequality and growth has no implications for whether a causal relationship exists between trade policy and growth. As a result, well over 90 different variables have been proposed as potential growth determinants (Durlauf and Quah 1999), each of which has some ex ante plausibility. As there are at best about 120 countries available for analysis in cross-sections (the number may be far smaller as a result of missing observations on some covariates), it is far from obvious how to formulate firm inferences about any particular explanation of growth.

This issue of open-endedness has not been directly dealt with in the literature. Instead, a number of researchers have proposed ways to deal with the robust-

ness of variables in growth regressions. The basic idea of this approach is to identify a set of potential control variables for inclusion in equation 1 as elements of Z_i. Inclusion of a variable in the final choice of Z_i requires that its associated coefficient prove to be robust with respect to the inclusion of other variables. Levine and Renelt (1992) introduced this idea to the growth literature, employing Edward Leamer's ideas on extreme bounds analysis (see Leamer 1983 and Leamer and Leonard 1983). In extreme bounds analysis, a coefficient is robust if the sign of its OLS estimate stays constant across a set of regressions representing different possible combinations of other variables. Sala-i-Martin (1997), arguing that extreme bounds analysis is likely to lead to the rejection of variables that do influence growth, proposes computing likelihood-weighted significance levels of coefficients across alternative regressions.

These proposals for dealing with the plethora of growth theories are useful, but neither is definitive as a way to evaluate model robustness.[3] The reason is simple. In these approaches a given coefficient will prove not to be robust if its associated variable is highly collinear with variables suggested by other candidate growth theories. This is obvious for the Sala-i-Martin approach, because collinearity affects significance levels. It is also true for extreme bounds analysis, in the sense that a given coefficient is likely to be highly unstable when alternative collinear regressors are included alongside its corresponding regressor. Hence these procedures will give sensible answers only if lack of collinearity is a "natural" property for a regressor that causally influences growth. Yet when one thinks about theories of how various causal determinants of growth are themselves determined, it is clear that collinearity is a property that one might expect to hold for important causal determinants of growth.[4] This is easiest to see by considering a recursive model for growth. Suppose that growth is causally determined by a single regressor, d_i, and that this regressor in turn depends causally on a third regressor, c_i, so that

(2)
$$g_i = d_i \gamma_d + \varepsilon_i$$
$$d_i = c_i \pi_c + \eta_i.$$

It is easy to construct cases (which will depend on the covariance structure of c_i, ε_i, and η_i) in which adding c_i to the growth equation will render d_i fragile.

Important recent papers by Doppelhofer, Miller, and Sala-i-Martin (2000) and Fernandez, Ley, and Steel (2001) have proposed ways to deal with regressor

---

3. Leamer's work on model uncertainty falls into two parts: a powerful demonstration of the importance of accounting for such uncertainty in making empirical claims, and a specific suggestion, extreme bounds analysis, for determining when regressors are fragile. The first constitutes a fundamental set of ideas. The second is a particular way of instantiating Leamer's deep ideas of accounting for model uncertainty and is more easily subjected to criticism. By analogy, Rawl's controversial use of minimax arguments to infer what rules are just in a society does not diminish the importance of his idea of the veil of ignorance. Economists have inappropriately used criticisms of extreme bounds analysis to ignore the conceptual issues raised by Leamer's work.

4. Leamer is quite clear on this point. See Leamer (1978, p. 172) for further discussion.

choice and hence at least indirectly with model open-endedness through the use of Bayesian model averaging techniques. We exploit the approach used in those papers and therefore defer discussion of them until section V.

Parameter Heterogeneity

A second problem with conventional growth analyses is the assumption of parameter homogeneity. The vast majority of empirical growth studies assume that the parameters that describe growth are identical across countries. This assumption is surely implausible. Does it really make sense to believe that a change in the level of a civil liberties index has the same effect on growth in the United States as in the Russian Federation? Although the use of panel data approaches to growth has addressed one aspect of this problem by allowing for fixed effects (Evans [1998] is particularly clear on this point), it has not addressed this more general question.

In some sense this criticism might seem unfair, as it presumably applies to any socioeconomic data set. After all, economic theory does not imply that individual units ought to be characterized by the same behavioral functions. That said, any empirical analysis necessarily will require a set of interpretable statistical properties that are common across observations; when homogeneity assumptions are or are not to be made is a matter of judgment. Our contention is that the assumption of parameter homogeneity is particularly inappropriate in studying complex heterogeneous objects, such as countries. See Draper (1997) for a general discussion of these issues.

Evidence of parameter heterogeneity has been developed in different contexts, such as in Canova (1999); Desdoigts (1999); Durlauf and Johnson (1995); Durlauf, Kourtellos, and Minkin (2000); Kourtellos (2000); and Pritchett (2000). These studies use very different statistical methods, but each suggests that the assumption of a single linear statistical growth model that applies to all countries is incorrect.[5] Put differently, the reporting of conditional predictive densities based on the assumption that all countries obey a common linear model may understate the uncertainty present when the data are generated by a family of models; Draper (1997) provides further analysis of this idea.

There has been substantial interest in the empirical growth literature in incorporating forms of parameter heterogeneity when panel data are available. Islam (1995) is an early analysis that allows constant terms to differ across country growth processes for a panel in which growth is measured in five-year intervals. In what appears to be the richest analysis of parameter heterogeneity to date, Lee, Pesaran, and Smith (1997) show how to allow for parameter heterogeneity for regressor slope parameters for a growth model employing annual data.

---

5. Conventional growth analyses give some attention to parameter heterogeneity between rich and poor countries: Barro (1996), for example, allows the effects of democracy on growth to differ between rich and poor countries.

The idea that panel data may be used to model rich forms of parameter heterogeneity is of course important; a comprehensive analysis is Pesaran and Smith (1995). However, this approach is of limited use in empirical growth contexts, because variation in the time dimension is typically small. This occurs for two reasons. First, many of the variables used as proxies for new growth theories do not vary over high frequencies. For some variables, such as political regime, this is true by their nature; for others, this is due to measurement. In any event, this means that cross-section variation must be used to uncover parameters. Second, there is a conceptual question of the appropriate time horizon over which to employ a growth model. High-frequency data will contain business cycle factors that are presumably irrelevant for long-run output movements. Hence it is difficult to see how annual or biannual data, for example, can be interpreted in terms of growth theories. In our view the use of long run averages has a powerful justification for identifying growth as opposed to cyclical factors.

Causality versus Correlation

A final source of skepticism about conventional growth empirics relates to a problem endemic to all structural inference in social science—the question of causality versus correlation. Many of the standard variables used to explain growth patterns—democracy, trade openness, rule of law, social capital, and the like—are as much outcomes of socioeconomic decisions and interrelationships as growth itself is. Hence there is an a priori case that the use of OLS estimates of the relationship between growth and such variables cannot be treated as structural any more than coefficients produced by OLS regressions of price on quantity can be. Yet the majority of empirical growth studies treat the various growth controls as exogenous variables and so rely on ordinary or heteroskedasticity-corrected least squares estimation. What is particularly ironic about the lack of attention to endogeneity is that it was precisely this lack of attention in early business cycle models that helped drive the development of rational expectations econometrics.

Recent econometric practice in growth has begun to employ instrumental variables to control for regressor endogeneity. This is particularly common for panel data sets where temporally lagged variables are treated as legitimate instruments. However, this trend toward using instrumental variables estimation has not satisfactorily addressed this problem. The reason is that the failure to account properly for the open-endedness of growth theories has important implications for the validity of instrumental variables methods.

What we mean by this is the following. For a regression of the form

$$(3) \qquad y_i = R_i\gamma + \varepsilon_i$$

the use of some set of instrumental variables I_i as instruments for R_i requires, of course, that each element of I_i be uncorrelated with ε_i. In the growth literature this is not a condition typically employed to motivate the choice of instruments. Instead, instruments are typically chosen exclusively because they are

in some sense exogenous, which operationally means that they are predetermined with respect to ε_i. Predetermined variables, however, are not necessarily valid instruments.

As discussed in Durlauf (2000), a good example of this pitfall can be found in Frankel and Romer (1996), which studies the relationship between trade and growth. Frankel and Romer argue that because trade openness is clearly endogenous, it is necessary to instrument the trade openness variable in a cross-country regression to consistently estimate the trade openness coefficient. To do this, they use a geographic variable, area, as an instrument and argue in favor of its validity that area is predetermined with respect to growth. Their argument that the instrument is predetermined is certainly persuasive. Nevertheless, it is hard to make an argument that it is a valid instrument. Is it plausible that country land size is uncorrelated with the omitted growth factors in their regression? The history and geography literatures are replete with theories of how geography affects political regime, development, and so on. For example, larger countries may be more likely to be ethnically heterogeneous, leading to attendant social problems. Alternatively, larger countries may have higher per capita military expenditures, which means relatively greater shares of unproductive government investment, higher distortionary taxes, or both. Our argument is not that any one of these links is necessarily empirically salient, but that the use of land area as an instrument presupposes the assumption that the correlations between land size and all omitted growth determinants are in total negligible. It is difficult to see how such an assumption can be defended when these omitted growth determinants are neither specified nor evaluated.

It is interesting to contrast the difficulties of identifying valid instruments in growth contexts with the relative ease with which this is done in rational expectations contexts. The reason for this difference is that rational expectations models are typically closed in the sense that a particular theory will imply that some combination of variables is a martingale difference with respect to some sequence of information sets. For the purposes of data analysis, rational expectations models therefore generate instrumental variables, that is, any variables observable at the time expectations are formed, whose orthogonality to expectation errors may be exploited to achieve parameter identification. Of course, rational expectation models can be faulted for imposing sufficiently wide-ranging restrictions on the economic behavior under study that some of the assumptions necessary for identification are not plausible; that is, for being insufficiently open-ended in the sense we have described. So the problems associated with theory open-endedness in growth are hardly nonexistent in other contexts.

III. Exchangeability

Inferences from any statistical model can only be made, of course, conditional on various prior assumptions that translate the data under study into a particu-

lar mathematical structure. One way to evaluate the plausibility of inferences drawn from empirical growth regressions is by assessing the plausibility of the assumptions made in making this translation. In the empirical growth literature it is easy to find examples where the assumptions employed to construct statistical models are clearly untenable. For example, researchers typically assume that the errors in a cross-section regression are jointly uncorrelated and orthogonal to the model's regressors.[6] Do they really wish to argue that no omitted factors exist that induce correlation across the innovations in the growth regressions associated with the model? More generally, it is easy to see that parameter heterogeneity and omitted variables, which, we argued in the previous section, are endemic to growth regressions, can each lead to a violation of the error uncorrelatedness assumption, the regressor orthogonality assumption, or both.

On the other hand, econometrics has a long tradition of identifying minimal sets of conditions under which coefficients and standard errors may be consistently estimated. Examples include the emphasis on orthogonality conditions between regressors and errors as the basis for OLS consistency (rather than the interpretation of the OLS estimators as the maximum likelihood estimates for a linear model with nonstochastic regressors and *i.i.d.* normal errors) or the use of mixing conditions to characterize when central limit theorems apply to dependent data (rather than the modeling of the series as a known autoregressive moving average process). Hence any critique of cross-country growth analyses that is based on the plausibility of particular statistical assumptions needs to argue that the violations of the assumptions invalidate the objectives of a given exercise.

In this section we argue that of the three econometric issues we have raised, the first two may be interpreted as examples of deviations of empirical growth models from a statistical "ideal" that allows for the sorts of inferences researchers wish to make in growth contexts. Our purpose is to establish a baseline for statistical growth models such that if a model does not meet this standard, a researcher needs to determine whether the reasons for this invalidate the goal of the empirical exercise. Hence the baseline does not describe a necessary requirement for empirical work, but instead helps define a strategy that we think empirical workers should follow in formulating growth models. When a model does not meet this standard, researchers should be prepared to argue that the violations of the standard do not invalidate the empirical claims they wish to make. This standard is based on a concept in probability known as *exchangeability*.[7]

6. In the subsequent discussion, we focus on OLS estimation of growth regressions. In the empirical growth literature examples can be found of heteroskedasticity corrections to relax assumptions of identical residual variances and instrumental variables to deal with violations of error/regressor orthogonality. Our discussion is qualitatively unaffected by either of these alternatives to OLS.

7. Bernardo and Smith (1994) provide a complete introduction to exchangeability. Draper and others (1993) develop a detailed argument on the importance of exchangeability to statistical inference. Our analysis is much indebted to their perspective.

Basic Ideas

A formal definition of exchangeability is as follows.

DEFINITION: EXCHANGEABILITY. A sequence of random variables η_i is exchangeable if, for every finite collection $\eta_1 \ldots \eta_K$ of elements of the sequence,

$$(4) \qquad \mu(\eta_1 = a_1, \ldots, \eta_K = a_K) =$$
$$\mu(\eta_{\rho(1)} = a_1, \ldots, \eta_{\rho(K)} = a_K)^8$$

where $\rho(\, . \,)$ is any operator that permutes the K indices.

Exchangeability is typically treated as a property of the unconditional probabilities of random variables. In regression contexts, however, it is often more natural to think in terms of the properties of random variables conditional on some information set. For example, in a regression, one is interested in the properties of the errors conditional on the regressors. We therefore introduce a second concept, *F*-conditional exchangeability.[9]

DEFINITION: *F*-CONDITIONAL EXCHANGEABILITY. For a sequence of random variables η_i and a collection of associated random vectors F_i, η_i is *F*-conditionally exchangeable if, for every finite collection $\eta_1 \ldots \eta_K$ of elements of the sequence,

$$\mu(\eta_1 = a_1, \ldots, \eta_K = a_K \mid \underset{\sim}{F}) =$$
$$\mu(\eta_{\rho(1)} = a_1, \ldots, \eta_{\rho(K)} = a_K \mid \underset{\sim}{F})$$

where $\rho(\, \cdot \,)$ is any operator that permutes the K indices and $\underset{\sim}{F} = \{F_1 \ldots F_K\}$.

If $F_i = \phi \; \forall \; i$, the empty set, *F*-conditional exchangeability reduces to exchangeability.

Associated with exchangeability and *F*-conditional exchangeability is the idea of partial exchangeability.

DEFINITION: PARTIAL EXCHANGEABILITY. A sequence of random variables η_i is partially exchangeable with respect to a sequence of random vectors Y_i if, for every finite collection $\eta_1 \ldots \eta_K$ of elements of the sequence,

$$(6) \qquad \mu(\eta_1 = a_1, \ldots, \eta_K = a_K \mid Y_i = \bar{Y} \; \forall \; i \in \{1 \ldots K\}) =$$
$$\mu(\eta_{\rho(1)} = a_1, \ldots, \eta_{\rho(K)} = a_K \mid Y_i = \bar{Y} \; \forall \; i \in \{1 \ldots K\})$$

where $\rho(\, \cdot \,)$ is any operator that permutes the K indices.

The key difference between exchangeability and partial exchangeability is the

---

8. Throughout, $\mu(\, . \,)$ is used to denote probability measures.

9. *F*-conditional exchangeability was originally defined in Kallenberg (1982). Ivanoff and Weber (1996) provide additional discussion. The notion of *F*-conditional exchangeability is rarely employed in the statistics literature and is not mentioned in standard textbooks such as Bernardo and Smith (1994). We believe the reason for this is that exchangeability analyses in the statistics literature generally focus on whether the units under study are exchangeable, rather than whether they are conditional on certain characteristics, the more natural notion in economic contexts.

conditioning on common values of some random vectors Y_is associated with the η_is in the partial exchangeability case. If Y_i is a discrete variable, partial exchangeability implies that a sequence may be decomposed into a finite or countable number of exchangeable subsequences.

Even though *F*-conditional exchangeability of model errors constitutes a stronger assumption than is needed for many of the interpretations of OLS, this exchangeability condition is nevertheless useful as a benchmark in the construction and assessment of statistical models. We make this claim for two reasons. First, this exchangeability concept helps organize discussions of the plausibility of the invariance of conditional moments that lie at the heart of policy relevant predictive exercises. Draper (1987, p. 458) describes the critical role of exchangeability in any predictive exercise:

> Predictive modeling is the process of expressing one's beliefs about how the past and future are connected. These connections are established through *exchangeability judgments*: with what aspects of past experience will the future be more or less interchangeable, after conditioning on relevant factors? It is not possible to avoid making such judgments; the only issue is whether to make them explicitly or implicitly.

Put in the context of growth analysis, the use of cross-country data to predict the behavior of individual countries presupposes certain symmetry judgments about the countries, judgments that are made precise by forms of exchangeability.

Second, exchangeability is separately important because of its implications for the appropriate statistical theory to apply in growth contexts. The reason for this relates to a deep result in probability theory known as de Finetti's Representation Theorem.[10] This theorem, formally stated in the technical appendix, establishes that the sample path of a sequence of exchangeable random variables behaves *as if* the random variables were generated by a mixture of *i.i.d.* processes. For empirical practice, de Finetti's Representation Theorem is important because it creates a link between a researcher's prior beliefs about the nature of the data under analysis (specifically, the properties of regression errors) that permits the researcher to interpret OLS estimates and associated test statistics in the usual way.[11]

10. See Bernardo and Smith (1994, chs. 4 and 6) for an insightful discussion of the nature and implications of the theorem and Aldous (1983) for a comprehensive mathematical development of various forms of the theorem.

11. Caution is needed in using de Finetti's theorem to calculate the distributions of regression estimators. For linear regression models of the form of equation 1, with normally distributed errors and nonstochastic regressors, Arnold (1979, p. 194) shows that "many optimal procedures for the model with *i.i.d.* errors are also optimal procedures for the model with exchangeably distributed errors . . . in the univariate case the best linear unbiased estimator and the ordinary least squares estimator are equal . . . as long as the experimenter is only interested in hypotheses about (the slope coefficients of the regression) he may act as though the errors were *i.i.d.*" Further, if the errors are non-normal, de Finetti's theorem leads one to expect analogous asymptotic equivalences. Similarly, we believe that analogies to de Finetti's theorem can be developed for stochastic regressors and *F*-conditional exchangeability, although as far as we know no such results have been established.

Exchangeability and Growth

How does exchangeability relate to the assumptions underlying cross-country growth regressions? These models typically assume that once the included growth variables in the model are realized, no basis exists for distinguishing the probabilities of various permutations of residual components in country-level growth rates, that is, these residuals are F-conditionally exchangeable, where F is the modeler's information set. Notice that F may include variables beyond those included in a growth regression as well as knowledge about nonlinearities or parameter heterogeneity in the growth process.

Various forms of exchangeability appear, in our reading of the empirical growth literature, to implicitly underlie many of the regression specifications. An implicit (F-conditional) exchangeability assumption is made whenever the empirical implementation of the growth trajectory for a single country from a given theoretical model is turned into a cross-country regression (typically after linearizing) by allowing the trajectory's state variables to differ across countries and appending an error term. Such an assumption of exchangeability has substantive implications for how a researcher thinks about the relationship between a given observation and others in a data set. Suppose that a researcher is considering the effect of a change in trade openness on a country, for example, Tanzania, in Sub-Saharan Africa. How does the researcher employ estimates of the effects of trade on growth in other countries to make this assessment? The answer depends on the extent to which the causal relationship between trade and growth in Tanzania can be uncovered using data from other countries.

More generally, notice how a number of modeling assumptions that are standard in conventional growth exercises are conceptually related to the assumptions that the model errors ε_i are F-conditionally exchangeable and that the growth rates g_i are partially exchangeable with respect to available information. Specifically,

1. The assumption that a given regression embodies all of a researcher's knowledge of the growth process is related to the assumption that the errors in a growth regression are F-conditionally exchangeable.
2. The assumption that the parameters in a growth regression are constant is related to the assumption that country-level growth rates are partially exchangeable.
3. The justification for the use of ordinary (or heteroskedasticity-corrected) least squares, as is standard in the empirical growth literature, is related to the assumption that the errors in a growth regression are exchangeable (or are exchangeable after a heteroskedasticity correction).

Our general claim is that exchangeability, in particular, F-conditional exchangeability of model errors, is an "incredible" (Sims 1980) assumption in the context of the standard cross-country regressions of the growth literature. (By a

standard regression, we refer to equation 1, in which a small number of regressors are assumed to explain cross-country growth patterns.[12]) For exchangeability to hold for a given regression and information set, the likelihood of a positive error for a given country—say, Japan—would need to be the same as that for any other country in the sample. In turn, for this to be true, no prior information could exist about the countries under study that would render the distribution of the associated growth residuals for these countries sensitive to permutations.

To repeat, exchangeability is not necessary to justify the estimation methods and structural interpretations conventionally given to cross-country growth regressions.[13] Hence our use of the term *related* in the three points above. What exchangeability does is provide a baseline, based on economic theory and a researcher's prior knowledge of the growth process, by which to assess cross-country regressions. Exchangeability is a valuable baseline for two reasons. First, the conditions under which various types of exchangeability do or do not hold for growth rates or model residuals can be linked to a researcher's substantive understanding of the growth process in ways that alternative sets of (purely statistical) assumptions on errors usually cannot be. In turn, once exchangeability is believed to be violated, a researcher can naturally link the reasons that exchangeability fails to hold to the question of whether the estimation methods used in the growth literature nevertheless can be expected to yield consistent parameter estimates and standard errors.

Second, exchangeability is important because it shifts the focus of specification analysis away from the question of theory inclusion (determining which variables need to be included in a growth regression to cover relevant structural growth determinants) to the identification of groups of countries that obey a common regression surface and hence can provide information on the growth process. This shift of emphasis is important for two reasons. First, for many growth determinants, the variables used to proxy for theories are very poor measures. For example, in the standard Gastil index of political rights, often used to measure levels of democracy, South Africa is ranked as high as or even higher than (depending on the period) the Republic of Korea for the period 1972–84. It is difficult to know what this means (political rights for whom?) and in what sense this rank ordering is relevant for the aspects of democracy conducive to growth. A more fruitful exercise is to identify groups of countries that obey a common, parsimonious growth model. Put differently, if, as seems plausible, many growth determinants such as political regime are common background variables for subsets of countries, a more productive empirical strategy may be to identify these subsets rather than to use crude empirical proxies for regime. Second, to the extent that nonquantifiable factors, such as "culture" (see Landes

---

12. To be fair, empirical growth papers often check the robustness of variables relative to a small number of alternative controls, but such robustness checks do not address exchangeability per se.

13. In section VI we return to the question of when the full force of exchangeability is useful for policy analysis.

2000), matter for growth, the identification of partially exchangeable subsets may be necessary for any sort of growth inferences.

To see how exchangeability plays a role in the leap from the identification of statistical patterns to structural inference, suppose that one runs the baseline Solow regression and observes that regression errors for the countries in Sub-Saharan Africa are predominately negative (as is the case). How does one interpret this finding? One can either attribute the finding to chance (the errors are, after all, zero mean with nonnegligible variance) or conclude that there was something about those countries that was not captured by the model. Easterly and Levine (1997), for example, develop a comprehensive argument on the role of ethnic divisions as a causal determinant of growth working from this initial fact. Or, put differently, Easterly and Levine (1997), from prior knowledge about the politics and cultures of these countries, developed their analysis on the basis that the Solow errors were not exchangeable, that is, that there was something about Sub-Saharan African countries that should have been incorporated into the Solow model.

Does the requirement of exchangeability imply the impossibility of structural inference whenever observational data are being studied? This would grossly exaggerate the import of our critique. Exchangeability of errors is conceivable for a wide range of models with observational data sets. For example, exchangeability seems to be a plausible assumption for statistical models based on the use of individual-level data sets, such as the Panel Study of Income Dynamics (PSID), once relevant information about the individuals under study is controlled for. One reason for this relates to the units of analysis. A basic difference between microeconomic data sets of this type and macroeconomic data sets of the type used in growth analysis is that macroeconomic observations pertain to large heterogeneous aggregates for which a great deal of information is known; information that can imply that exchangeability does not hold. In addition, the large size of individual-level data sets such as the PSID means that the range of possible control variables is much greater than that for growth. By this we mean something deeper than "the more data points, the more regressors may be included." Instead, we argue that large data sets of the type found in microeconomics will contain observations on groups of individuals who are sufficiently similar with respect to observables that they may be plausibly regarded as representing exchangeable observations.

That said, we fully accept that exchangeability for observations on objects as complicated as countries may well be problematic. Will our knowledge of the histories and cultures of the countries in cross-country regressions ever be embedded in the regressions to such an extent that the exchangeability requirement is met? This question is at the heart of many of the controversies about the empirical growth literature.

To summarize, conventional growth econometrics has failed to consider the ways in which appropriate exchangeability concepts may or may not hold for

the specific models analyzed. This failure in turn renders these studies difficult if not impossible to interpret, because one must know whether any exchangeability violations that are present invalidate the statistical exercise being conducted. We therefore concur with Draper and others (1993, p. 1), who argue that

> statistical methods are concerned with combining information from different observational units and with making inferences from the resulting summaries to prospective measurements on the same or other units. These operations will be useful only when the units to be combined are judged to be *similar* (comparable or homogeneous) . . . judgments of similarity involve concepts more primitive than probability, and these judgments are central to preliminary activities that all statisticians must perform, even though probability specifications are absent or contrived at such a preliminary stage.

Exchangeability and Causality

Though exchangeability is a useful benchmark for understanding some of the major sources of skepticism about growth regressions, it does not bear in any obvious way on the third of our general criticisms, the lack of attention to causality versus correlation in growth analysis. For example, following a nice example due to Goldberger (1991), a regression of parental height on daughter height can have a perfectly well-defined set of exchangeable errors, so that parental heights are partially exchangeable, yet the interpretation of the associated regression coefficient is obviously noncausal. More generally, causality is a different sort of question than the other issues we have addressed, in that it cannot be reduced to a question of whether the data fulfill a generic statistical property. As Heckman (2000, p. 89) notes, "causality is a property of a model . . . many models may explain the same data and . . . assumptions must be made to identify causal or structural models." And (2000, p. 91):

> Some of the disagreement that arises in interpreting a given body of data is intrinsic to the field of economics because of the conditional nature of causal knowledge. The information in any body of data is usually too weak to eliminate competing causal explanations of the same phenomenon. There is no mechanical algorithm for producing a set of "assumption free" facts or causal estimates based on those facts.

In our subsequent discussion we do not address strategies for dealing with questions of causality. Instead, we focus on model uncertainty, which presupposes that causality uncertainty within a given model has been addressed by suitable assumptions by the analyst. In doing this, we are not diminishing the importance of thinking about causal inference; instead, we believe that causal arguments require judgments about economic theory and qualitative information about the problem at hand that represent issues separate from those we address.

IV. A Digression on Noneconometric Evidence

Regression analyses of the type conventionally done are useful mechanisms for summarizing data and uncovering patterns. These techniques are not, as currently employed, particularly credible ways to engage in causal inference. Before proceeding to econometric alternatives, we wish to point out the importance of integrating different sources of information in the assessment of growth theories. These sources are often the basis on which exchangeability can be questioned in a particular context.

The economic history literature is replete with studies that are of enormous importance in adjudicating different growth explanations, yet this literature usually receives only lip service in the growth literature.[14] An exemplar of historical studies that can speak to growth debates is Clark (1987), which explores the sources of productivity differences between cotton textile workers in New England and those workers in other countries in 1910. These differences were immense—a typical New England textile worker was about six times as productive as his counterpart in China or India and more than twice as productive as his counterpart in Germany. Clark painstakingly shows that these differences cannot be attributed to differences in technology, education, or management.[15] Instead, they seem to reflect cultural differences in work and effort norms. Such studies have important implications for understanding why technology may not diffuse internationally and how poverty traps may emerge, and should play a far greater role in the empirical growth literature.

Historical and qualitative studies also play a crucial role in the development of credible statistical analyses. One reason for this is that these sorts of studies provide information on the plausibility of identifying assumptions that are made to establish causality. Further, our discussion on exchangeability and growth analysis may be interpreted as arguing that a researcher needs to do one of two things to claim that a regression provides causal information. The researcher must make a plausible argument that, given the many plausible growth theories and plausible heterogeneity in the way different causal growth factors affect different countries, the errors in a particular growth regression are nevertheless exchangeable. Or, the researcher can make the argument that the violations of exchangeability in the regression occur in ways that do not affect the interpretation of the coefficients and standard errors from those that are employed. To some extent exchangeability judgments must be made prior to a statistical exercise, as Draper and others (1993) note above. Where does information of this type come from? Often from qualitative and historical work. Hence, the detailed study of individual countries that is a hallmark of work by the World Bank, for example, plays an invaluable role in allowing credible statistical analysis.

---

14. There are notable exceptions, such as Easterly and Levine (1997) and Prescott (1998).

15. See also Wolcott and Clark (1999), which provides detailed evidence that managerial differences cannot explain the low productivity in Indian textiles.

V. Modeling Model Uncertainty

The main themes of our criticisms of current econometric practice may be summarized as two claims:

1. The observations in cross-country growth regressions do not obey various exchangeability assumptions given the information available on the countries under study.

This implies that:

2. Model uncertainty is not appropriately incorporated into empirical growth analyses.

There are no panaceas for the interpretation problems we have described for growth regressions. Although our formulation of model uncertainty can reduce the dependence of empirical growth studies on untenable exchangeability or other assumptions, growth regressions will always rely on untestable and possibly controversial assumptions if causal or structural inferences are to be made. It may be impossible, for example, to place every possible growth theory in a common statistical analysis, so critiques based on theory open-endedness will apply, at some level, to our own suggestions. Further, we will not be able to model all aspects of uncertainty about partial exchangeability of growth rates. However, we do not regard this as a damning defect. Empirical work always relies on judgment as well as formal procedures, what Draper and others (1993, p. 16) refer to as "the role of leaps of faith" in constructing statistical models. What we wish to do is reduce the number and magnitude of such leaps.

General Framework

We assume that the structural growth process for country i obeys a linear structure that applies to all countries j that are members of class $J(i)$. Suppose that this model is described by a set of regressors S that we partition into a subset X and a scalar z. Our analysis focuses on how to employ data to uncover β_z, the coefficient that determines the effect of z_i on country i's growth. We work with models of the form:

$$(7) \qquad g_j = S_j \zeta + \varepsilon_j = X_j \pi + z_j \beta_z + \varepsilon_j, \, j \in J(i).$$

When a given model represents the "true" or correct specification of the growth process for countries in J, the sequence of residuals ε_j will be F-exchangeable. The information set F comprises the total available information to a researcher about the countries. For our purposes, F will consist of a collection of regressors available to a researcher; S is a subset of these. The idea that a model consists of the specification of a set of growth determinants, (S_j), and the specification of a set of countries with common parameters, $J(i)$, that together render the associated model errors F-conditionally exchangeable, will, as we shall see, parallel our earlier discussion of the first two sources of criticisms of growth regressions.

It is skepticism about the claim that a particular model is correctly specified in the sense we have described that renders many of the empirical claims in the growth literature not credible.

The standard approach to statistical analysis in the growth literature can be thought of as using a single model M and given data set D to analyze model parameters. Suppose that the goal of the exercise is to uncover information about a particular parameter β_z. From a frequentist perspective, this involves calculating an estimate of the parameter β_z along with an associated standard error for the estimate. From a Bayesian perspective, this involves calculating the posterior density $\mu(\beta_z \mid D,M)$. We will employ the Bayesian framework in our subsequent discussion. That said, we will be interested in relating our analysis to frequentist analyses of growth. For this reason we shall often employ a "leading case" in the analysis. As described in the technical appendix, under some conditions the posterior mean of the set of regression coefficients in equation 7 equals the OLS estimates of the parameters and the posterior variance/covariance matrix equals the variance/covariance matrix of the OLS estimates. We will use this equivalence repeatedly in the next section.

Formulating Types of Model Uncertainty

Suppose that there exists a universe of models, M with typical element M_m, that are possible candidates for the "true" growth model that generated the data under study; the true model is assumed to lie in this set.[16] This universe is generated from two types of uncertainty. First, there is *theory uncertainty*. In particular, we assume that there is a set X of possible regressors to include in a growth regression whose elements correspond to alternative causal growth mechanisms. In our framework a theory is defined as a particular choice of regressors for a model of the form of equation 7. Second, there is *heterogeneity uncertainty*. By this we mean that there is uncertainty as to which countries make up $J(i)$, that is are partially exchangeable with country i.[17] In the presence of these types of uncertainty a researcher will be interested not in $\mu(\beta_z \mid D,M_m)$ for a particular M_m but in $\mu(\beta_z \mid D)$; the exception, of course, is when the correct growth theory and the set of countries that are partially exchangeable with country i are known with certainty to the modeler.

This dichotomy of model uncertainty can, at least in principle, incorporate other forms of uncertainty as well. Consider the question of nonlinearities in the growth process. One could attempt to deal with *functional form uncertainty* through the addition of regressors. Examples would include adding regressors that are nonlinear functions of the initial set of theory-based regressors (appealing to Taylor series-type or other approximations) or adding regressors whose

---

16. It is possible to consider contexts where no model in M is correct, as discussed in Bernardo and Smith (1994), but that is beyond the scope of this paper.

17. These two types of uncertainty are not independent; for example, theory uncertainty may induce heterogeneity uncertainty.

values are zero below some threshold and equal to a theory-based regressor above that threshold (as suggested by such models as Azariadis and Drazen [1990]. In this sense heterogeneity uncertainty is no different from theory uncertainty.

It is possible to integrate theory uncertainty and some forms of heterogeneity uncertainty into a common variable selection framework. Doing so has the important advantage that it allows us to draw on new developments in the statistics literature stemming from an important paper by Raftery, Madigan, and Hoeting (1997). By definition, theory uncertainty is a question of variable inclusion. To see how to interpret heterogeneity uncertainty in a similar way, we proceed as follows. For a given regressor set S, suppose that one believes that the countries under study may be divided into two subsets with associated subscripts A_1 and A_2 such that the countries within each subset are partially exchangeable, but that countries in one subset may not be partially exchangeable with countries in the other because of parameter heterogeneity. Each of these subsets is characterized by a linear equation so that

$$(8) \qquad g_j = X_j\pi + z_j\beta_z + \varepsilon_j \text{ if } j \in A_1$$

and

$$(9) \qquad g_j = X_j\pi' + z_j\beta'_z + \varepsilon_j \text{ if } j \in A_2.$$

This last equation can be rewritten as

$$(10) \qquad g_j = X_j\pi + z_j\beta_z + X_j (\pi' - \pi) + z_j (\beta'_z - \beta_z) + \varepsilon_j \text{ if } j \in A_2.$$

Therefore, the two equations can be combined into a single growth regression of the form:

$$(11) \quad g_j = X_j\pi + z_j\beta_z + X_j \delta_{j,A_2} (\pi' - \pi) + z_j \delta_{j,A_2} (\beta'_z - \beta_z) + \varepsilon_j, \text{ if } j \in A_1 \cup A_2$$

where $\delta_{j,A_2} = 1$ if $j \in A_2$, 0 otherwise. The additional regressors $X_j\delta_{j,A_2}$ and $z_j\delta_{j,A_2}$ therefore produce a common regression for all observations.[18] Of course, this type of procedure is often done in empirical work; our purpose in this development here is to emphasize how heterogeneity uncertainty may be explicitly modeled in terms of variable inclusion. Notice that it is straightforward to generalize this procedure to multiple groups of partially exchangeable countries. This procedure is not completely general in that it restricts the sort of possible parameter heterogeneity allowed; for example, each country is not allowed a separate set of coefficients. To allow for this more general type of heterogeneity would require moving to an alternative structure, such as a hierarchical linear model (see Schervish 1995, ch. 8); we plan to pursue this in subsequent work.

18. When heterogeneity uncertainty is introduced, the variable z will be associated with different parameters for different countries. For ease of exposition we let β_z refer to the relevant parameter for the country i that is of interest.

Posterior Probabilities

Once a researcher has formulated a space of possible models, it is relatively straightforward to calculate posterior probabilities that do not rely on the assumption that one model is true. In the presence of model uncertainty the calculation of $\mu(\beta_z \mid D)$ requires integrating out the dependence of the probability measure $\mu(\beta_z \mid D, M_m)$ on the model M_m. By Bayes's rule, the posterior density of a given coefficient conditional only on the observed data is

$$(12) \qquad \mu(\beta_z \mid D) = \sum_m \mu(\beta_z \mid D, M_m)\mu(M_m \mid D),$$

which can be rewritten as

$$(13) \qquad \mu(\beta_z \mid D) \propto \sum_m \mu(\beta_z \mid D, M_m)\mu(D \mid M_m)\mu(M_m),$$

where $\propto$ means "is proportional to," $\mu(D \mid M_m)$ is the likelihood of the data given model M_m, and $\mu(M_m)$ is the prior probability of model M_m. This formulation gives a way of eliminating the conditioning of the posterior density of a given parameter on a particular model choice.

Calculations of this type originally appeared in Leamer (1978) and are reported in Draper (1995). Leamer (1978, p. 118) gives the following derivations of the conditional mean and variance of β_z given the data D:

$$(14) \qquad E(\beta_z \mid D) = \sum_m \mu(M_m \mid D)E(\beta_z \mid D, M_m)$$

and

$$(15) \qquad var(\beta_z \mid D) = E(\beta_z^2 \mid D) - (E(\beta_z \mid D))^2 =$$
$$\sum_m \mu(M_m \mid D)(var(\beta_z \mid M_m, D) + (E(\beta_z \mid D, M_m))^2) - (E(\beta_z \mid D))^2 =$$
$$\sum_m \mu(M_m \mid D)var(\beta_z \mid M_m, D) + \sum_m \mu(M_m \mid D)(E(\beta_z \mid D, M_m) - (E(\beta_z \mid D))^2.$$

As discussed in Leamer (1978) and Draper (1995), the overall variance of the parameter estimate β_z depends on the variance of the within-model estimates (the first term in equation 15) and the variance of the estimates across models (the second term in equation 15).

Equation 12 and the related expressions are all examples of Bayesian model averages. The methodology surrounding Bayesian model averaging is specifically developed for linear models with uncertainty about variable inclusion in Raftery, Madigan, and Hoeting (1997).[19] Doppelhofer, Miller, and Sala-i-Martin (2000), focusing on theory uncertainty only, compute a number of measures of variable robustness based on the application of this formula to growth regressions and conclude that initial income is the "most robust" regressor. Fernandez, Ley, and Steel (1999) also employ Bayesian model averaging for theory uncertainty, focusing on the explicit computation of posterior coefficient distributions. Our own development should be read as an endorsement and extension of the analyses in

---

19. The survey by Hoeting and others (1999) provides a nice introduction to model averaging techniques. See also Wasserman (2000).

these articles. Our formulation differs in two respects from previous work. First, we treat heterogeneity uncertainty as well as theory uncertainty as part of overall model uncertainty. Draper and others (1993) provide a general overview of the importance of accounting for heterogeneity uncertainty in constructing credible empirical exercises. As our discussion illustrates, heterogeneity uncertainty can be treated as a question of variable inclusion, so the ideas in Doppelhofer, Miller, and Sala-i-Martin (2000) and Fernandez, Ley, and Steel (1999) can be extended to this domain in a straightforward fashion. Second, we develop an explicit decision-theoretic approach to interpreting growth regressions. As far as we are aware, this analysis is new.

Outliers

One important concern in the empirical growth literature has revolved around the role of outliers in determining various empirical claims. A famous example is the role of the Botswana observation in determining the estimated magnitude of social returns to equipment investment (DeLong and Summers 1991, 1994; and Auerbach, Hassett, and Oliner 1994).[20]

Outliers can be dealt with in a straightforward fashion. There are three strategies one can pursue. First, one can always employ a within-model estimator that is designed to be robust to outliers. As Temple (2000) points out, one can employ a trimmed least square estimator (one that drops or downweights observations whose associated OLS residuals are large) in estimating each model's parameters and still employ whatever posterior analysis one wishes. Second, one can explicitly allow the density for model errors to accommodate outliers. For example, one can model errors as drawn from a mixture distribution. Third, and most promising in our view, one can employ a Bayesian bagging procedure due to Clyde and Lee (2000). Bayesian bagging ("bagging" is an abbreviation for bootstrap aggregating) was introduced by Breiman (1996) to improve the performance of what he called "unstable" prediction and modeling methods. A method is "unstable" when small changes in the data set lead to large changes in the method's output. Intuitively, the Clyde and Lee procedure constructs bootstrap data sets from the empirical distribution function of a data set, computes a model average for each sample, and then averages these results. (See their article for details.) Clyde and Lee provide reasons to think that this modification of model averaging will be robust to outliers.

That said, the ex post analysis of outliers, as was carried out in the Botswana case we described, is often problematic; as Leamer (1978, p. 265) remarks, the mechanical and typically ad hoc dropping of outliers both leads to invalid statistical conclusions and ignores valuable information.

20. The role of outliers in growth regressions has been somewhat overstated; for example, the DeLong and Summers results are far more robust to the inclusion or exclusion of Botswana than is often asserted, as a careful reading of DeLong and Summers (1991, 1994) and Auerbach, Hassett, and Offner (1994) clearly reveals. Temple (1998) is a more persuasive example of the importance of dealing with outliers.

Priors on the Space of Possible Models

An important issue in the implementation of the model averaging approach that we describe is the choice of the prior distribution on the space of models. For the problem of variable inclusion, this is typically handled by assuming that all 2^k possible models (where k is the number of regressors that may be placed in a given model) have equal probability; Fernandez, Ley, and Steel (1999) follow this procedure in their analysis. The procedure in essence assumes that the prior probability that a given regressor is in a model is 1/2. Doppelhofer, Miller, and Sala-i-Martin (2000) make the alternative assumption that for a regression whose expected number of included regressors is $\bar{k}$, the probability of inclusion of a given regressor is $\bar{k} / k$. They make this assumption to avoid "a very strong prior belief that the number of included variables should be large" (2000, pp. 15–16).

These alternative approaches to setting model priors are not very appealing from the perspective of economic theory. Clearly, the addition of a given regressor to the set of possible regressors should affect the probabilities with which other variables are included. It is unclear, for example, why the effect of ethnicity on growth should be independent of the effect of democracy, as it can easily be imagined that one will affect growth only if the other does as well. The conventional approaches to modeling the space of priors ignore this fact.

This problem is closely associated with a standard criticism of the "irrelevance of independent alternatives" assumption in choice theory, originally due to Debreu (1960) and later instantiated in the choice literature as the "red bus/blue bus problem" (see Ben-Akiva and Lerman 1985, sec. 3.7). In discrete choice theory irrelevance of independent alternatives means that the ratio of choice probabilities between any two alternatives should be unaffected by the presence of a third. As pointed out by Debreu, this assumption is untenable if the third choice is a close substitute for one of the other two. For the analysis of growth regressions, the priors we have discussed suffer from a similar problem, although the reasons are more complicated. As noted above, the likelihood that one growth theory matters may covary positively with whether another one matters. Further, because the variables employed to capture growth theories are often crude proxies for underlying theories, their inclusion probabilities could covary positively, as each helps measure some common growth determinants. For example, contra Doppelhofer, Miller, and Sala-i-Martin (2000, n. 15), the likelihood that political assassinations predict growth differences could be positively associated with the likelihood that revolutions predict growth differences, as each helps instrument the unobservable variable "political instability."

We have no advice to offer on how to deal with this problem, because its resolution will depend on one's priors on the space of underlying growth models, as determined by the interconnections between particular growth theories. In our view, it makes more sense at this stage of development to treat the prior distribution over models as a benchmark for reporting posterior statistics. (A number of Bayesians have developed a similar view of priors; see, for example, the dis-

cussion of "robust" priors in Berger 1987, p. 111.) Because the complexity of the growth process speaks to the strong likelihood that a large number of growth factors substantively matter, the uniform prior of Fernandez, Ley, and Steel (1999) makes the most sense at this stage in providing a benchmark.

That said, there is nothing theoretically compelling about the assumption that the inclusion probability of each regressor is 1/2. We therefore believe that it might make sense in future work to report values for some benchmark alternative probabilities, in order to help evaluate the robustness of results. By choosing inclusion probabilities lower than 1/2, it is possible to incorporate the spirit of the Doppelhofer, Miller, and Sala-i-Martin concerns without having to form prior beliefs on the expected number of regressors in a model, which seems extremely problematic.

VI. Toward a Policy-Relevant Growth Econometrics

The framework developed in the previous section provides a general way of describing model uncertainty in growth regressions. It does not, however, provide any guidance on how to determine what variables should be included in a regression, or on when to regard the sign or magnitude of a regression coefficient as robust. The reason is that the posterior densities embodied in equations 12 to 15 are nothing more than data summaries. As such, they can inform policy analysis only to the extent that they are integrated with a specific formulation of the decision problem of a policymaker. Hence it is necessary to develop an explicit decision-theoretic basis for assessing growth data. The decision-theoretic framework we describe explicitly incorporates the various forms of model uncertainty associated with possible violations of exchangeability, as discussed in the previous section. In this section we discuss the use of growth regressions to inform empirical analysis when one of the growth controls is under the control of a policymaker. Many of the purported policy variables included in growth regressions—for example, indices of political stability—are not necessarily tightly linked to the variables over which a policymaker has control. The framework we describe can be generalized to incorporate a more complicated relationship between growth determinants and policy than the one we analyze here.

The decision-theoretic perspective involves moving away from a specific concern with a particular hypothesis to an evaluation of the implications of a given set of data for a particular course of action. Kadane and Dickey (1980, p. 247) argue

> The important question in practice is not whether a true effect is zero, for it is already known not to do exactly zero, but rather, How large is the effect? But then this question is only relevant in terms of How large is important? This question in turn depends on the use to which the inference will be put, namely, the utility function of the concerned scientist. Approaches which attempt to explain model specification from the viewpoint of the inappro-

priate question, Is it true that . . . ? have a common thread in that they all proceed without reference to the utility function of the scientist. And therefore, from the decision theory point of view, they all impose normative conditions on the utility function which are seldom explicit and often far from the case.

Substituting *policymaker* for *scientist* in this quotation makes it clear why policy-relevant growth econometrics needs to explicitly integrate policy objectives and empirical practice. Our approach is well summarized by Kass and Raftery (1994, p. 784): "The decision making problem is solved by maximizing the posterior expected utility of each course of action considered. The latter is equal to a weighted average of the posterior expected utilities conditional on each of the models, with the weights equal to the posterior model probabilities." In other words, we argue that policy-relevant econometrics needs explicitly to identify the objectives of the policymaker and then calculate the expected consequences of a policy change.

Policy Assessment: Basic Ideas

The basic posterior coefficient density described by equation 12 and the associated first and second moments described by equations 14 and 15 represent data summaries and as such have no implications for either inference or policy assessment. The goal of a policy analysis is not to construct such summaries but to assess the consequences of changes in a policy. Similarly, such data summaries do not imply the validity of particular rules for data evaluation or inference. For example, the assessment of whether regressors are robust, such as is in extreme bounds analysis or the comparison of models using Bayes factors,[21] may not be appropriate for certain policy exercises. Put differently, decisions on whether to treat regressors as robust and the like should, for the purposes of policy analysis, be derived from the policymaker's assessment of the expected payoffs associated with alternative policies.

In this section we explore policy assessment when model uncertainty has been explicitly accounted for. The purpose of this exercise is twofold. First, it captures what we believe is the appropriate way for policymakers to draw inferences from data. Second, it shows that various rules for the assessment of regressor fragility, such as extreme bounds analysis, will arise in such exercises. A critical feature of this approach to model assessment is it illustrates that the evaluation of regressor robustness can be derived from particular aspects of the policymaker's objective function.

For expositional purposes we initially suppose that the goal of an empirical exercise is to evaluate the effect of a change dz_i[22] in some scalar variable that is

---

21. For any two models M_m and $M_{m'}$, the Bayes factor $B_{m,m'}$ is defined as $\mu(D \mid M_m) / \mu(D \mid M_{m'})$. Kass and Raftery (1994) provide an extensive overview of the use of Bayes factors in model evaluation.

22. Without loss of generality, we generally assume that $dz_i > 0$.

under the control of a policymaker and believed to have some effect on growth. Therefore, the decisionmaker's set of actions A is $\{0, dz_i\}$. This decision rule is based on a vector observable data D. This means that a decisionmaker chooses a rule $\phi(\cdot)$ that maps D to A so that

(16)
$$\phi(D) = dz_i \text{ if } D \in D_1$$
$$\phi(D) = 0 \text{ otherwise.}$$

D_1 is therefore the acceptance region for the policy change. We assume that the "true" linear growth model is a causal relationship that will allow evaluation of the effect of this change.

Because we restrict ourselves to linear models, the analysis of the policy decision is particularly straightforward, as $\mu(\beta_z \mid D)$ will describe the posterior distribution of the effect of a marginal change in z on growth in a given country. A marginal policy intervention in country i can be evaluated as follows. Let z_i denote the level of a policy instrument in country i. This instrument appears as one of the regressors in the linear model that describes cross-country growth. Suppose one has the option of either keeping the policy instrument at its current value or changing it by a fixed amount dz_i. Let g_i denote the growth rate in the country in the absence of the policy change, and $g_i + \beta_z dz_i$ the growth rate with the change. Finally, let $V(g_i, O_i)$ denote the utility value of the growth rate to the policymaker. O_i is a placeholder vector that contains any factors relating to country i that affect the policymaker's utility.

An expected utility assessment of the policy change can be based on the comparison

(17)
$$E(V(g_i + \beta_z dz_i, O_i) \mid D) - E(V(g_i, O_i) \mid D).$$

Calculations of the expected utility differential in equation 17 implicitly contain all information relevant to a policy assessment. From the perspective of policy evaluation, the various rules that have been proposed for the assessment of regressor robustness should be an implication of this calculation. Notice that this calculation requires explicitly accounting for model uncertainty, because the conditioning is always done solely with respect to the data.

Policy Assessment under Alternative Utility Functions

In this section we consider the implications of some alternative utility functions for the analysis of growth regressions. Our goal is to show how particular utility functions will lead a policymaker to decide whether or not to implement a policy on the basis of aspects of the posterior distribution of β_z. We do not claim that the utility functions we examine are particularly compelling. We have chosen them to illustrate what sort of utility functions can justify some of the standard ways of interpreting growth regressions.

RISK NEUTRALITY. Suppose that V is linear and increasing in the level of growth, that is,

(18) $$V(g_i,O_i) = \alpha_0 + \alpha_1 g_i, \alpha_1 > 0.$$

For this policymaker the relevant statistic is the posterior mean of the regressor coefficient. In this case it is straightforward to see that the policy change is justified if the expected value of the change in the growth rate is positive, that is,

(19) $$\sum_m \mu(M_m \mid D)E(\beta_z \mid D,M_m) > 0.$$

When the prior model probabilities are equal, this is equivalent to the condition

(20) $$\sum_m \mu(D \mid M_m)E(\beta_z \mid D,M_m) > 0$$

so the likelihoods $\mu(D \mid M_m)$ determine the relative model weights.

MEAN/VARIANCE UTILITY OVER CHANGES IN THE GROWTH RATE. Suppose that a policymaker has preferences that relate solely to changes in the growth rate, as opposed to its level. The idea here is that a policymaker assesses a policy relative to the baseline g_i. Operationally, we assume that one chooses the elements of O_i and the functional form of $V(\cdot, \cdot)$ so that

(21) $$E(V(g_i + \beta_z dz_i,O_i) \mid D) - E(V(g_i, O_i) \mid D) =$$
$$\alpha_0 E(\beta_z dz_i \mid D) + \alpha_1 var(\beta_z dz_i \mid D)^{1/2}, \alpha_0 > 0, \alpha_1 < 0.$$

When $|\alpha_0/\alpha_1| = 1/2$, this utility specification implies that the policymaker will act only if the t-statistic (the posterior expected value of β_z divided by its posterior standard deviation) is greater than 2. Hence this specification, at least qualitatively, corresponds to the standard econometric practice of ignoring regressors whose associated t-statistics are less than 2.

From a decision-theoretic perspective, the conventional practice of ignoring "statistically insignificant" coefficients (by which we mean coefficients whose posterior standard errors are more than twice their posterior expected values) can be justified only in very special cases. First, it is necessary to assume that the form of risk aversion of the policymaker applies to the standard deviation rather than to the variance of the change in growth. Otherwise, the desirability of the policy will depend on the magnitude of dz_i. For example, if the utility function is

(22) $$E(V(g_i + \beta_z dz_i,O_i) \mid D) - E(V(g_i, O_i) \mid D) =$$
$$\alpha_0 E(\beta_z dz_i \mid D) + \alpha_1 var(\beta_z dz_i \mid D), \alpha_0 > 0, \alpha_1 < 0$$

with $|\alpha_0/\alpha_1| = 1/2$, there will be a threshold level T such that for all $0 < dz_i \le T$ a policy change increases the policymaker's utility.[23] Therefore, the rule of ignoring regressors with t-statistics less than 2 presupposes a very specific assumption about how risk affects the policymaker's utility. Second, if equation 22 is the correct utility function, the policymaker may still choose to act with the fixed dz_i level we started with under (conventionally defined) statistical insignificance

___

23. This is an example of the famous result of Pratt that one will always accept a small amount of a fair bet. We plan to address the question of the optimal choice of dz_i in future work.

or, alternatively, may decline to act when the coefficient is statistically significant. These possibilities can be generated through appropriate choices of $|\alpha_0/\alpha_1|$.

KNIGHTIAN UNCERTAINTY AND MAXIMIN PREFERENCES. In the examples we have studied thus far we have allowed all uncertainty about the correct model M_m to be reflected in the posterior model probabilities $\mu(M_m \mid D)$. An alternative approach to model uncertainty, one in the tradition of Knightian uncertainty, assumes that an additional layer of uncertainty exists in the environment under study that may be interpreted as a distinct type of risk, sometimes called ambiguity aversion, as will be seen below.

As before, let M denote the universe of possible growth models. A risk sensitive utility function for the policymaker can be defined as

$$(23) \qquad (1- e)E(V(g_i,O_i) \mid D) + e(inf_{M_m \in M}E(V(g_i,O_i) \mid D,M_m)).$$

In this equation e denotes the degree of ambiguity.

This equation is motivated by recent efforts to reconceptualize utility theory in light of results such as the classic Ellsberg paradoxes. For example, if experimental subjects are given a choice between (1) receiving \$1 if they draw a red ball at random from an urn that they know contains 50 red balls and 50 black balls and (2) receiving \$1 if they draw a red ball when the only information available is that the urn contains 100 red and black balls, the subjects typically choose the first, "unambiguous" urn (Camerer 1995, p. 646). Clearly, if subjects were Bayesians who placed a flat prior on the distribution of the balls in case 2, they would be indifferent between the two options.[24]

Experimental evidence of ambiguity aversion has led researchers—including Anderson, Hansen, and Sargent (1999); Epstein and Wang (1994); and Gilboa and Schmeidler (1989)—to consider formal representations of preferences that exhibit ambiguity aversion. One popular representation, studied in Epstein and Wang (1994), replaces expected utility calculations of the form $\int u(\omega)dP(\omega)$ with $inf_{p \in P} \int u(\omega)dP(\omega)$, where P is a space of possible probability measures. When this space contains a single element, this second expression reduces to the first, which is the standard expected utility formulation. A variant of this formulation is to assume that P consists of a set of mixture distributions $(1 - e)P_0 + eP_1$, where P_0 is a baseline probability measure that a policymaker believes to be true, P_1 is the least favorable of all possible probability measures for the policymaker, and e represents the strength of the possibility that this measure applies. When the universe of alternative processes for growth is the space of linear models that we have described, one can replace P with M_m and P with M and obtain the second part of equation 23. In using this specification, we do not claim that it is the only sensible way to model ambiguity aversion by policymakers. We introduce

24. See Camerer (1995) for additional examples of ambiguity aversion as well as a survey of the implications of different results in the experimental economics literature for utility theory.

it to illustrate how recent developments in decision theory may be linked to econometric practice.

We can explore the effects of this additional uncertainty on our analysis by considering the two specifications of V studied above. First, assume that V is linear and increasing, while equation 23 characterizes the ambiguity aversion we have described. In this case the policy change dz is justified if

$$(24) \qquad (1 - e)E(\beta_z \mid D) + e(inf_{M_m \in M}E(\beta_z \mid D, M_m)) > 0.$$

When $e = 1$, the policy action will be taken only when $E(\beta_z \mid D, M_m) > 0$ for all $M_m \in M$. This has an interesting link to OLS coefficients for different models in M. In the leading case described in the technical appendix, the posterior expectation $E(\beta_z \mid D, M_m)$ equals $\hat{\beta}_{z,m}$, the OLS coefficient associated with the regressor z for model M_m. If $e = 1$, this utility function would then mean that a policymaker will choose to implement dz_i if the OLS coefficient estimate of β_z is positive for every model in M.

Alternatively, assume that the policymaker is risk-averse in the sense that equation 21 describes his utility function. In this case the policy change should be implemented if

$$(25) \qquad (1 - e)(\alpha_0 E(\beta_z \mid D) + \alpha_1 var(\beta_z \mid D)^{\frac{1}{2}}) + \\ e(inf_{M_m \in M}(\alpha_0 E(\beta_z \mid D, M_m) + \alpha_1 var(\beta_z \mid D, M_m)^{\frac{1}{2}})) > 0.$$

Again, this rule has an interesting link to OLS parameter estimates. If $e = 1$ and $|\alpha_0/\alpha_1| = 1/2$, then for the leading case in the technical appendix, the policymaker will not act unless the OLS regression coefficient $\hat{\beta}_{z,m}$ is positive and statistically significant (in the sense that the t-statistic is at least 2) for each model in M. (Here, we rely on the additional fact that for the leading case, $var(\beta_z \mid D, M_m)$ equals the OLS variance of $\hat{\beta}_{z,m}$.)

The policy rules that hold for $e = 1$ are closely related to the recommendations made by Leamer for assessing coefficient fragility through extreme bounds analysis (see Leamer 1983 and Leamer and Leonard 1983). In extreme bounds analysis, recall that when a regressor "flips signs" across specifications, this is argued to imply that the regressor is fragile. From the perspective of policy recommendations, we interpret this notion of fragility to mean that no policy change should be made when there is a model of the world under which the policy change can be expected to make things worse off. This suggests that extreme bounds analysis is based on a maximin assumption of some type. Our derivations show that this intuition can be formalized.

This derivation of extreme value analysis appears to complement a number of the objections raised against it by Granger and Uhlig (1990) and McAleer, Pagan, and Volker (1985). Both these articles argue that extreme bounds analysis can lead to spurious rejections of regressors as a result of changes in sign induced by regressions that are, by standard tests, misspecified. In our view these criticisms need to be developed from the perspective of the objectives of the empirical exercise. Put differently, the salience of these critiques of extreme

bounds analysis requires that one reject the utility functions we have described as supporting extreme bounds analysis.

Further, we believe that our derivations provide an appropriate way of modifying extreme bounds analysis—through the use of utility functions, such as equation 23 for $0 < e < 1$. For such cases the relative goodness of fit of different models will be relevant to the empirical exercise. As is well known (see Wasserman 2000, p. 94 for a nice exposition), when Bayesian model selection between two models is based on posterior odds ratios and the prior odds on the models are equal, the posterior odds ratios will equal the ratio of their likelihoods, that is, the posterior odds will reflect the relative likelihoods of the data under the alternative models. Further, as the amount of data becomes large enough, for this special case of equal prior odds, the model with the minimum Kullback-Leibler Information Criterion (KLIC) distance to the "truth" will be revealed. If the set of models under scrutiny includes the true model, the true model will be revealed in large samples.[25] Thus in our context, under our assumption of equal prior odds across models, we may expect the data in large samples to ultimately place greater weight on models whose KLIC distance is closer to the true model.[26]

By choosing $0 < e < 1$ for policymaker utility functions such as equations 24 and 25, one can retain the ambiguity aversion that justifies (in a limiting case) extreme bounds analysis. In particular, one can reflect a policymaker's desire to avoid harm when he faces scientific ambiguity, but at the same time prevent him from being so ambiguity-averse that he fails to take welfare-enhancing actions that are supported by relatively good posterior odds under available scientific evidence (especially when samples are large enough to contain *some* policy-relevant predictive information). Notice that our treatment avoids the criticism of Bayes factors in Kadane and Dickey (1980) that the weights do not account for the purpose of the empirical exercise.

ALTERNATIVE UTILITY FUNCTIONS. In the previous section we assumed that the policymaker cares only about the level of or change in growth induced by a policy change. It is of course possible to imagine other plausible utility functions for a policymaker. One possibility is to assume that a policymaker evaluates a utility on the basis of changes in the expected value of growth within a regime; formally, one assumes that there exists a function ψ such that

$$(26) \qquad \psi(E(g_i + \beta_z dz_i \mid D, M_m) - E(g_i \mid D, M_m))$$

---

25. See White (1994) for a discussion of measures of closeness based on KLIC and how various estimators achieve minimum KLIC distance to the true mode in large samples.

26. There are some subtleties involved in making the argument we have sketched precise. For example, regularity conditions need to be assumed to justify assertions about the relationship between KLIC distance minimization and quasi-maximum likelihood estimation. Furthermore, as Fernandez, Ley, and Steel (2001) point out, the form of priors for parameters *within* models raises thorny issues. Nevertheless, we believe that this heuristic argument is useful.

measures the utility for a policy change conditional on a particular growth model. Again assuming the leading case where $E(\beta_z \mid D, M_m)$ equals the OLS coefficient $\hat{\beta}_{z,m}$, linearity of the expected growth process implies that the expected utility from dz_i, once one accounts for model uncertainty, is

$$
\begin{aligned}
(27) \qquad E(\psi(E(\beta_z dz_i \mid D, M_m)) \mid D) &= \sum_m \mu(M_m \mid D)\psi(E(\beta_z dz_i \mid D, M_m)) \\
&= \sum_m \mu(M_m \mid D)\psi(\hat{\beta}_{z,m} dz_i)
\end{aligned}
$$

When $\psi(\cdot)$ is linear, this reduces to the risk-neutral case discussed earlier. However, alternative utility functions can produce very different decision rules. For example, suppose that either $\psi(c) = -\infty$ if $c < 0$, $\psi(\cdot)$ bounded otherwise, or $\psi(c) = -\infty$ if $c > 0$, $\psi(\cdot)$ bounded otherwise. One will then have the implied decision rule that a single sign change in the OLS coefficient estimate $\hat{\beta}_{z,m}$ as one moves across models is sufficient to imply that the policymaker should not act to either increase or decrease z_i by dz. This type of utility function induces behavior mimicking that found under Knightian uncertainty.

At first glance this might appear to be an unreasonable utility function for a policymaker. This conclusion is at least partially incorrect. Suppose that each state of the world is indexed by the growth process that is "true" under it. The utility of the policymaker will then depend on both the growth rate that is expected to prevail and the state of the world under which it transpires. For example, suppose that there is a model of the world in which the expected effect of democracy on growth is negative. Such a model could be one whose features imply that a policymaker is particularly wary of reducing growth by changing a given policy instrument. For example, if there is a (positive probability) model of the world in which democracy is especially fragile and may not survive a growth reduction, a policymaker might be especially wary of the policy change for fear this would prove to be the correct model of the world.

This type of argument can be formalized by considering model-dependent utility specifications. Suppose that conditional on model M_m, the utility from a policy change is equal to

$$
(28) \qquad U(E(\beta_z \mid D, M_m)dz_i, M_m) = U(\hat{\beta}_{z,m} dz_i, M_m)
$$

so that the posterior expected utility of the policy change is

$$
(29) \qquad E(U(\hat{\beta}_{z,m} dz_i, M_m) \mid D) = \sum_m \mu(M_m \mid D)U(\hat{\beta}_{z,m} dz_i, M_m).
$$

Manipulating $U(\cdot, \cdot)$, one can produce (under the leading case) a result that is consistent with refusing to act whenever the posterior mean $\hat{\beta}_{z,m}$ is negative for at least one M_m, thereby producing extreme bounds–like behavior in the sense that one would not choose $dz_i > 0$, even though for all other models $\hat{\beta}_{z,m}$ is positive.

Policy Analysis and Exchangeability

A decision-theoretic approach of the type we have advocated makes clear the importance of a growth model being rich enough for a researcher to plausibly

regard the observations as *F*-conditionally exchangeable. Suppose that a researcher is using data from I countries to provide a recommendation for the optimal choice of z_i subject to some constraint set Z_i for country i. In other words, a researcher is attempting to solve the problem

(30) $$max_{z_i \in Z_i} E(V(g_i, z_i))$$

where information in computing this expression is taken from the regression described by equation 1.

What information in equation 1 is relevant to this calculation? The answer depends on the shape of *V*. Suppose that *V* is linear in growth rates, as in equation 18 above. The only information needed about the growth process as described by equation 1 is the posterior expected value of β_z. In our leading case, the OLS coefficient in a growth regression will be sufficient for policy analysis as long as all countries are described by a common linear model. Growth rates need not be partially exchangeable, because partial exchangeability requires symmetry with respect to all moments of the growth process. Similarly, suppose that *V* is quadratic. In this case one will need only the second moments from the posterior densities generated by equation 1 to apply to country i; partial exchangeability is still not necessary.

However, if *V* is arbitrary, one will need to employ equation 1 to obtain information on the full conditional distribution $F(\varepsilon_i \mid X_i, Z_i)$. To reveal this type of statistic from cross-country data, one will require full *F*-conditional exchangeability of the type we have discussed.

VII. An Empirical Example

In this section we reconsider an important growth study, Easterly and Levine (1997), which examines the role of ethnic conflict in growth.[27] We chose to reexamine this study for three reasons. The study is widely regarded as quite important in the growth literature. It has important implications for policy and the sorts of advice and advocacy an international organization would engage in. And the authors of the study have done an admirable job of making their data and programs publicly available.

Easterly and Levine's analysis is designed to explain why in standard cross-country growth regressions the performance of Sub-Saharan Africa[28] is so much worse than that of the rest of the world. Rather than remain content with modeling this phenomenon as a fixed effect (a dummy variable) for these countries, Easterly and Levine argue that a major cause of the poor growth performance is the presence of ethnic conflict in these countries. They construct a measure of ethnic diversity to proxy for this conflict. This variable is substantially larger for

---

27. We thank Duncan Thomas for suggesting to us that the findings in Easterly and Levine (1997) warrant reexamination.

28. See the data appendix for the list of the countries in Sub-Saharan Africa.

Sub-Saharan Africa than for the rest of the world. Inclusion of the variable in a cross-country growth regression reduces the size of the African fixed effect and is itself statistically significant. Easterly and Levine (1997, p. 1241) conclude that "the results lend support to the theories that interest group polarization leads to rent-seeking behavior and reduces the consensus for public goods, creating long-run growth tragedies."

Our reexamination of this study has an explicitly narrow focus. In our view it is important to see whether and how the influence of ethnolinguistic heterogeneity on growth depends on what other variables are included in the regression. Further, a natural alternative to the claim that the African growth experience is different because of an omitted variable, ethnolinguistic heterogeneity, is that other growth determinants influence Africa differently than they do the rest of the world. Put differently, parameter heterogeneity is a natural alternative explanation.

We therefore conduct the following analysis to account for the effect of model uncertainty on Easterly and Levine's results. A data appendix describes the variables we employ; these are identical to those used in Easterly and Levine (1997). The data are based on decade-long average observations for the 1960s, 1970s, and 1980s, except where indicated in the appendix. We focus on a reexamination of Easterly and Levine's equation 3, table IV, which by conventional measures (such as the statistical significance of all included variables) is arguably their strongest regression in support of the role of ethnic diversity in growth. Our results using this regression are reported in column 1 of our table 1. The key variable of interest is ELF60, a measure of ethnic diversity in each country in 1960.

We explore the role of model uncertainty in two ways. We first consider the impact of theory uncertainty on inferences about the determinants of growth. We do this by constructing a universe of models that consists of all possible combinations of the variables in Easterly and Levine's baseline regression. This exercise should be interpreted as a robustness check for Easterly and Levine's results. To perform this exercise, we employ an approximation algorithm whereby posterior model probabilities are replaced with their maximum likelihood estimates. We perform the subsequent calculation of the posterior mean and standard deviation of each regression coefficient using formulas 14 and 15.[29]

Our results incorporating theory uncertainty are reported in column 2 of table 1. Interestingly, we find that the evidence of a role for ethnic diversity in the growth process is slightly strengthened through the model averaging technique. Specifically, the posterior mean of ELF60 is –0.02 under model averaging, compared with the –0.017 estimate reported by Easterly and Levine. Our primary conclusion from this exercise is that Easterly and Levine's main result is robust to theory uncertainty as we have characterized it.

29. See the computational appendix for details on the calculation of these quantities. Ethnolinguistic heterogeneity is not, of course, directly subject to a policymaker's control, so we do not explore the issues raised in section VI. The policy importance of the variable stems from the implications of its importance to questions of institutional design.

TABLE 1. OLS and Bayesian Model Averaging Coefficient Estimates and
Standard Errors Using Data from Easterly and Levine (1997)

	[1]	[2]	[3]	[4]	[5]	[6]
Intercept term	—	—	—	—	0.4013	0.1382
	—	—	—	—	(0.3985)	(0.0336)
Dummy for Sub-Saharan Africa	–0.0113	–0.0031	0.9558	0.0761	—	—
	(0.0048)	(0.0053)	(0.3704)	(0.0302)	—	—
Dummy for Latin America and the Caribbean	–0.0191	–0.0197	–0.0197	–0.0184	—	—
	(0.0036)	(0.0042)	(0.0035)	(0.0037)	—	—
Dummy for 1960s	–0.2657	–0.2200	–0.3643	–0.0028	—	—
	(0.0998)	(0.1765)	(0.1328)	(0.0326)	—	—
Dummy for 1970s	–0.2609	–0.2154	–0.3520	0.0009	0.0080	0.0050
	(0.0997)	(0.1745)	(0.1332)	(0.0325)	(0.0134)	(0.0079)
Dummy for 1980s	–0.2761	–0.2298	–0.3650	–0.0143	–0.0038	–0.0024
	(0.0996)	(0.1751)	(0.1336)	(0.0325)	(0.0132)	(0.0058)
Log of initial income	0.0870	0.0756	–0.1090	0.0218	–0.0696	–0.0004
	(0.0254)	(0.0444)	(0.0986)	(0.0088)	(0.1171)	(0.0027)
Log of initial income squared	–0.0063	–0.0056	0.0070	–0.0022	0.0044	–0.0000
	(0.0016)	(0.0029)	(0.0067)	(0.0006)	(0.0088)	(0.0002)
Log of schooling	0.0117	0.0130	–0.0220	0.0130	–0.0131	–0.0017
	(0.0042)	(0.0056)	(0.0216)	(0.0045)	(0.0194)	(0.0077)
Assassinations	–12.8169	–3.3629	–377.3810	–30.6120	–306.4870	–343.4434
	(9.2709)	(7.8137)	(165.5661)	(86.9027)	(158.4484)	(181.6948)
Financial depth	0.0162	0.0111	0.1010	0.0129	0.0774	0.0104
	(0.0058)	(0.0083)	(0.0497)	(0.0075)	(0.0483)	(0.0278)
Black market premium	–0.0188	–0.0219	–0.0130	–0.0207	–0.0171	–0.0039
	(0.0045)	(0.0053)	(0.0098)	(0.0043)	(0.0107)	(0.0081)
Fiscal surplus/GDP	0.1210	0.1717	0.1200	0.1382	0.1654	0.0948
	(0.0314)	(0.0411)	(0.0874)	(0.0357)	(0.0986)	(0.1071)
Ethnic diversity (ELF60)	–0.0169	–0.0222	–0.2020	–0.1437	–0.1516	–0.1595
	(0.0060)	(0.0066)	(0.0376)	(0.0279)	(0.0353)	(0.0327)

[1] Ordinary least squares estimates for model "ALL".

[2] Bayesian model averaging estimates for model "ALL".

[3] Ordinary least squares estimates for model "ALL + ALL*I(AFRICA)"; composite coefficient estimates and standard errors reported. AFRICA, LATINCA, and DUM60 dropped from AFRICA-specific set of regressors.

[4] Bayesian model averaging estimates for model "ALL + ALL*I(AFRICA)"; composite coefficient estimates and standard errors reported. AFRICA, LATINCA, and DUM60 dropped from AFRICA-specific set of regressors.

[5] Ordinary least squares on AFRICA subsample.

[6] Bayesian model averaging on AFRICA subsample.

Note: Standard errors are in parentheses.

As we have emphasized, theory uncertainty is not the only form of model uncertainty that needs to be accounted for in cross-country analysis. We therefore next incorporate heterogeneity uncertainty. Following equation 11, we do this by constructing for each regressor x_i in the baseline regressors a corresponding variable $x_i\delta_{j,A_A}$, where $\delta_{j,A_A} = 1$ if country j is in Sub-Saharan Africa and 0 otherwise. This allows for the possibility that the Sub-Saharan African countries have different growth parameters than the rest of the world. Column 3 in table 1 re-

ports the OLS values and standard errors of the regressor coefficients for the African countries; column 4 reports the same statistics when model averaging is done over the augmented variable set. Column 5 reports OLS estimates of the growth regression coefficients and standard errors when the African subsample is analyzed in isolation; column 6 reports the corresponding model average results.

Our explorations of the role of heterogeneity uncertainty provide a rather different picture of the role of ethnicity in African growth than of its role in the rest of the world. The coefficient estimates for Africa are about 7–10 times greater than the corresponding estimates for the world.[30] This result is extremely striking and makes clear that the operation of ethnic heterogeneity on growth is different in Africa, not just the levels of ethnic heterogeneity. Further, a comparison of the other regressor coefficients for Africa with those of the rest of the world makes clear that the growth observations for African countries should not be treated as partially exchangeable with the growth rates of the rest of the world.

These results in no way diminish the importance of Easterly and Levine's findings. In fact, our exercises show that their basic claims are robust to a limited variable uncertainty exercise. Our finding of parameter heterogeneity with respect to ethnolinguistic heterogeneity suggests a direction along which to extend their research. Our results illustrate how additional insights can be obtained by explicitly controlling for model uncertainty.

Finally, we again note that this reexamination is quite narrow. A full-scale study should at a minimum include explicit calculations and presentation of the predictive distribution of the effects of the policy change on growth. Fernandez, Ley, and Steel (1999) provide a good illustration of how to present results of this type. More generally, the reporting of results should always include the information necessary to calculate the posterior expected utility changes of the policymaker. Our own reporting is useful for mean/variance utility functions, but not for the others we have discussed. In addition, we have not allowed for parameter heterogeneity for countries outside Sub-Saharan Africa; doing so is a natural extension of this exercise. The results we report should be treated as suggestive, in this sense, of more elaborate examinations of the role of ethnic heterogeneity in the growth process.

VIII. Conclusions

This paper has had two basic aims. First, we attempted to delineate the major criticisms of cross-country growth regressions and to show how to interpret two

---

30. Similar results are obtained when one compares Sub-Saharan Africa with the rest of the world. When the Sub-Saharan African countries are dropped from the data set, the OLS estimate for the ELF60 regressor is –0.0115 with an associated standard error of 0.006. The associated values when model averaging is done across different regressor combinations (to check for robustness to theory uncertainty) are –0.013 and 0.009. By conventional levels, one would conclude that ethnicity is marginally statistically significant outside Sub-Saharan Africa.

of these criticisms, theory uncertainty and parameter uncertainty, as violations of a particular assumption—*F*-conditional exchangeability—in the residual components of growth models. Second, we outlined a framework for conducting and interpreting growth regressions. For conducting regressions, we advocated an explicit modeling of theory and heterogeneity uncertainty and the use of model averaging to condition out strong assumptions. For interpreting regressions, we argued that the policy objectives associated with a given exercise must be made explicit in the analysis. We outlined a decision-theoretic approach to growth regressions and explored its relationship to conventional approaches to assessing model robustness. Finally, in an empirical application we showed how attention to model uncertainty can provide new insights into the relationship between ethnicity and growth.

To amplify some earlier remarks, we do not believe that there is a single privileged way to conduct statistical or, for that matter, empirical analysis in the social sciences. Persuasive empirical work always requires judgments and assumptions that cannot be falsified or confirmed within the statistical procedure being employed.[31] Indeed, this is the reason that we have not included a treatment of how to provide more robust arguments in favor of causality in this article. What we hope is that this article has provided some initial steps toward the development of a language in which policy-relevant empirical growth research may be better expressed.

Computational Appendix

All model averaging calculations were done using the program bicreg, which was written in SPLUS by Adrian Raftery and is available at *www.research.att.com/~volinsky/bma.html*. Given the large number of possible models, this program, as is standard in the model averaging literature, uses a search algorithm that explores only a subset of the model space; the key feature of the design of the algorithm is that it ensures that the search proceeds along directions such that it is likely to cover models that are relatively strongly supported by the data. We follow the procedure suggested by Madigan and Raftery (1994); see Raftery, Madigan, and Hoeting (1997); and Hoeting and others (1999) for additional discussion. Though the reader should see those papers for a full description of the search algorithm, Hoeting and others (1999, p. 385) provide a nice intuitive description:

> 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 $\text{pr}(M_0/D)/\text{pr}(M_1/D)$. 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.

31. See Draper and others (1993) and Mallows (1998) for valuable discussions of such issues.

In implementing the model averaging procedure, the algorithm we employ uses an approximation, due to Raftery (1995), based on the idea that, because for a large enough number of observations, the posterior coefficient distribution will be close to the maximum likelihood estimator, and so one can use the maximum likelihood estimates to avoid the need to specify a particular prior. We refer the reader to Raftery (1994) as well as to Tierney and Kadane (1986) for 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; an important recent contribution is Fernandez, Ley, and Steel (2001).

Data Appendix: Variable Definitions

All data are the same as those used in Easterly and Levine (1997).

- *AFRICA:* Dummy variable for Sub-Saharan African countries, as defined by the World Bank. These countries are Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Cape Verde, the Central African Republic, Chad, Comoros, Democratic Republic of Congo, Republic of Congo, Côte d'Ivoire, Djibouti, Equatorial Guinea, Ethiopia, Gabon, The Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Rwanda, São Tomé and Principe, Senegal, Seychelles, Sierre Leone, Somalia, South Africa, Sudan, Swaziland, Tanzania, Togo, Uganda, Zambia, and Zimbabwe.
- *ASSASS:* Number of assassinations per 1,000 population.
- *BLCK:* Black market premium, defined as log of 1 + decade average of black market premium.
- *DUM60:* Dummy variable for 1960s.
- *DUM70:* Dummy variable for 1970s.
- *DUM80:* Dummy variable for 1980s.
- *ELF60:* A measure of ethnic diversity, equalling an index of ethnolinguistic fractionalization in 1960. This variable measures the probability that two randomly selected individuals from a given country will not belong to the same ethnolinguistic group.
- *GYP:* Growth rate of real per capita GDP.
- *LATINCA:* Dummy variable for countries in Latin America and the Caribbean.
- *LLY:* Financial depth, measured as the ratio of liquid liabilities of the financial system 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.
- *LRGDP:* Log of real per capita GDP measured at the start of each decade.
- *LRGDPSQ:* Square of LRGDP.

- *LSCHOOL:* Log of 1 + average years of school attainment, quinquennial values (1960–65, 1970–75, 1980–85).
- *SURP:* Fiscal surplus/GDP: Decade average of ratio of central government surplus to GDP, both in local currency, local prices.

TECHNICAL APPENDIX

1. De Finetti's Representation Theorem

De Finetti's theorem establishes that the symmetry inherent in the concept of the exchangeability of errors leads to a representation of the joint distribution of the errors in terms of an integral of the joint product of identical marginal distributions against some conditional distribution function. The theorem is as follows.

If η_i is an infinite exchangeable sequence with associated probability measure P, there exists a probability measure Q over F, the space of all distribution functions on R, such that the joint distribution function $F(\eta_{i-j} \ldots \eta_i \ldots \eta_{i+k})$ for any finite collection $\eta_{i-j} \ldots \eta_i \ldots \eta_{i+k}$ may be written as

(A-1) $$F(\eta_{i-j} \ldots \eta_i \ldots \eta_{i+k}) = \int \prod_{r=-j}^{k} F(\eta_{i+r}) dQ(F).$$

See Bernardo and Smith (1994, p. 177), for this formulation of de Finetti's theorem as well as a proof.

2. Some Relations between OLS Estimates and Bayesian Posteriors

For the linear model

(A-2) $$g_i = S_i \zeta + \varepsilon_i \; i = 1 \ldots I$$

suppose that (1) conditional on $S_1 \ldots S_I$, the ε_is are independent and identically distributed and jointly normal; the marginal distribution of the typical element is $N(0, \sigma_\varepsilon^2)$, (2) σ_ε^2 is known, and (3) prior information on ζ is characterized by the noninformative (improper) prior

(A-3) $$\mu(\zeta) \propto c$$

where c is a constant. Denote the OLS estimate (as well as the classical maximum likelihood estimate) of ζ as $\hat{\zeta}$, and denote the data matrix of regressors in equation A-2 as S.

   As shown for example in Box and Tiao (1973, p. 115), the posterior density of the parameter vector ζ given the available data D, $\mu(\zeta \mid D, M)$, is, under our assumptions, multivariate normal. Specifically,

(A-4) $$\mu(\zeta \mid D) \sim N(\hat{\zeta}, (S'S)^{-1} \sigma_\varepsilon^2)$$

The posterior density of any particular coefficient can of course be calculated from this vector density. Under the assumptions justifying A-4, the posterior

mean and variance of ζ therefore correspond to the standard OLS estimates of the parameter vector and its associated covariance matrix.

When σ_ε^2 is unknown, the posterior density of ζ can also be characterized and related to OLS estimates. Formally, if σ_ε^2 is unknown and has a noninformative prior

$$(A\text{-}5) \qquad\qquad \mu(\sigma_\varepsilon^2) \propto \sigma^{-2},$$

then it can be shown (Box and Tiao 1973, p. 117) that

$$(A\text{-}6) \qquad\qquad \mu(\zeta \mid D, \sigma_\varepsilon^2) \sim N(\hat{\zeta}, (S'S)^{-1}\sigma_\varepsilon^2).$$

For reasonably large samples, σ_ε^2 can be replaced with the OLS estimate $\hat{\sigma}_\varepsilon^2$ so that, approximately,

$$(A\text{-}7) \qquad\qquad \mu(\zeta \mid D) \sim N(\hat{\zeta}, (S'S)^{-1}\hat{\sigma}_\varepsilon^2)$$

and again the posterior mean and variance of ζ may be equated with the corresponding OLS estimates. We refer to this as the "leading case" in the text.

In our evaluation of growth models, we have emphasized the role of *F*-exchangeable, as opposed to independent and identically distributed errors. De Finetti's theorem provides a link between exchangeability and independence and so motivates our use of this leading case.

REFERENCES

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Aldous, D. 1983. "Exchangeability and Related Topics." In *École d'Été de Probabilités de Saint Flour XIII*. Lecture Notes in Mathematics Series, no. 1117. New York: Springer-Verlag.

Anderson, E., L. Hansen, and T. Sargent. 1999. "Risk and Robustness in Equilibrium." Department of Economics, Stanford University. Processed.

Arnold, S. 1979. "Linear Models with Exchangeably Distributed Errors." *Journal of the American Statistical Association* 74:194–99.

Auerbach, A., K. Hassett, and S. Oliner. 1994. "Reassessing the Social Returns to Equipment Investment." *Quarterly Journal of Economics* 109:789–802.

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

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

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

Ben-Akiva, M., and S. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, Mass.: MIT Press.

Berger, J. 1987. *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag.

Bernard, A., and S. Durlauf. 1996. "Interpreting Tests of the Convergence Hypothesis." *Journal of Econometrics* 71:161–72.

Bernardo, J., and A. Smith. 1994. *Bayesian Theory*. New York: John Wiley and Sons.

Box, G., and G. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. New York: John Wiley and Sons. Reprinted 2000.

Breiman, L. 1996. "Bagging Predictors." *Machine Learning* 26:123–40.

Camerer, C. 1995. "Individual Decision Making." In J. Kagel and A. Roth, eds., *Handbook of Experimental Economics*. Princeton, N.J.: Princeton University Press.

Canova, F. 1999. "Testing for Convergence Clubs in Income Per-Capita: A Predictive Density Approach." Department of Economics, University of Pompeu Fabra, Spain. Processed.

Clark, G. 1987. "Why Isn't the Whole World Developed? Lessons from the Cotton Mills." *Journal of Economic History* 47:141–73.

Clyde, M., and H. Lee. 2000. "Bagging and the Bayesian Bootstrap." Duke University, Department of Statistics, Durham, N.C. Processed.

Debreu, G. 1960. "Review of R. D. Luce, *Individual Choice Behavior: A Theoretical Analysis.*" *American Economic Review* 50:186–88.

DeLong, J. B., and L. Summers. 1991. "Equipment Investment and Economic Growth." *Quarterly Journal of Economics* 106:445–502.

———. 1994. "Equipment Investment and Economic Growth: Reply." *Quarterly Journal of Economics* 109:803–7.

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

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 no. 7750, National Bureau of Economic Research, Cambridge, Mass.

Draper, D. 1987. "Comment: On the Exchangeability Judgments in Predictive Modeling and the Role of Data in Statistical Research." *Statistical Science* 2:454–61.

———. 1995. "Assessment and Propagation of Model Uncertainty." *Journal of the Royal Statistical Society*, Series B 57:45–70.

———. 1997. "On the Relationship between Model Uncertainty and Inferential/Predictive Uncertainty." School of Mathematical Sciences, University of Bath. Processed.

Draper, D., J. Hodges, C. Mallows, and D. Pregibon. 1993. "Exchangeability and Data Analysis." *Journal of the Royal Statistical Society*, Series A 156:9–28.

Durlauf, S. 2000. "Econometric Analysis and the Study of Economic Growth: A Skeptical Perspective." In R. Backhouse and A. Salanti, eds., *Macroeconomics and the Real World*. Oxford: Oxford University Press.

Durlauf, S., and P. Johnson. 1995. "Multiple Regimes and Cross-Country Growth Behavior." *Journal of Applied Econometrics* 10:365–84.

Durlauf, S., and D. Quah. 1999. "The New Empirics of Economic Growth." In J. Taylor and M. Woodford, eds., *Handbook of Macroeconomics*. Amsterdam: North Holland.

Durlauf, S., A. Kourtellos, and A. Minkin. 2000. "The Local Solow Growth Model." Forthcoming in the *European Economic Review, Papers and Proceedings*.

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

Epstein, L., and T. Wang. 1994. "Intertemporal Asset Pricing Behavior under Knightian Uncertainty." *Econometrica* 62:283–322.

Evans, P. 1998. "Using Panel Data to Evaluate Growth Theories." *International Economic Review* 39:295–306.

Fernandez, C., E. Ley, and M. Steel. 1999. "Model Uncertainty in Cross-Country Growth Regressions." Department of Economics, University of Edinburgh (also forthcoming, *Journal of Applied Econometrics*).

———. 2001. "Benchmark Priors for Bayesian Model Averaging." *Journal of Econometrics* 100:381–427.

Frankel, J., and D. Romer. 1996. "Trade and Growth: An Empirical Investigation." Working Paper 5476, National Bureau of Economic Research. Cambridge, Mass.

Freedman, D. 1991. "Statistical Models and Shoe Leather." In P. Marsden, ed., *Sociological Methodology 1991*. Cambridge: Basil Blackwell.

———. 1997. "From Association to Causation via Regression." In V. McKim and S. Turner, eds., *Causality in Crisis*. South Bend, Ind.: University of Notre Dame Press.

Galor, O. 1996. "Convergence? Inferences from Theoretical Models." *Economic Journal* 106:1056–69.

Gilboa, I., and David Schmeidler. 1989. "Maximin Expected Utility with Nonunique Prior." *Journal of Mathematical Economics* 18:141–53.

Goldberger, A. 1991. *A Course in Econometrics*. Cambridge, Mass.: Harvard University Press.

Granger, C., and H. Uhlig. 1990. "Reasonable Extreme-Bounds Analysis." *Journal of Econometrics* 44:159–70.

Heckman, J. 2000. "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective." *Quarterly Journal of Economics* 115:45–97.

Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky. 1999. "Bayesian Model Averaging: A Tutorial." *Statistical Science* 14:382–401.

Islam, N. 1995. "Growth Empirics: A Panel Data Approach." *Quarterly Journal of Economics* 110:1127–70.

Ivanoff, B., and N. Weber. 1996. "Some Characterizations of Partial Exchangeability." *Journal of the Australian Mathematical Society*, Series A 61:345–59.

Kadane, J., and J. Dickey. 1980. "Bayesian Decision Theory and the Simplification of Models." In J. Kmenta and J. Ramsey, eds., *Evaluation of Econometric Models*. New York: Academic Press.

Kallenberg, O. 1982. "Characterizations and Embedding Properties in Exchangeability." *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 60:249–8l.

Kass, R., and A. Raftery. 1994. "Bayes Factors." *Journal of the American Statistical Association* 90:773–95.

Kourtellos, A. 2000. "A Projection Pursuit Approach to Cross-Country Growth Data." Department of Economics, University of Wisconsin. Processed.

Landes, D. 2000. "Culture Makes Almost All the Difference." In L. Harrison and S. Huntington, eds., *Culture Matters*. New York: Basic Books.

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

———. 1983. "Let's Take the Con Out of Econometrics." *American Economic Review* 73:31–43.

Leamer, E., and H. Leonard. 1983. "Reporting the Fragility of Regression Estimates." *Review of Economics and Statistics* 65:306–17.

Lee, K., M. H. Pesaran, and R. Smith. 1997. "Growth and Convergence in a Multi-Country Stochastic Solow Model." *Journal of Applied Econometrics* 12:357–92.

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

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

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* 89:1535–46.

Mallows, C. 1998. "The Zeroth Problem." *American Statistician* 52:1–9.

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

McAleer, M., A. Pagan, and P. Volker. 1985. "What Will Take the Con Out of Econometrics?" *American Economic Review* 75:293–307.

Pack, H. 1994. "Endogenous Growth Theory: Intellectual Appeal and Empirical Shortcomings." *Journal of Economic Perspectives* 8:55–72.

Pesaran, M. H., and R. Smith. 1995. "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels." *Journal of Econometrics* 68:79–113.

Prescott, E. 1998. "Needed: A Theory of Total Factor Productivity." *International Economic Review* 39:525–52.

Pritchett, L. 2000. "Patterns of Economic Growth: Hills, Plateaus, Mountains, and Plains." *World Bank Economic Review* 14:221–50.

Quah, D. 1996a. "Convergence Empirics across Economies with Some Capital Mobility." *Journal of Economic Growth* 1:95–124.

———. 1996b. "Empirics for Growth and Economic Convergence." *European Economic Review* 40:1353–75.

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

Raftery, A. 1995. "Bayesian Model Selection in Social Research." In P. Marsden, ed., *Sociological Methodology 1995*. Cambridge: Blackwell.

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

Raiffa, H., and R. Schlaifer. 1961. *Applied Statistical Decision Theory*. New York: John Wiley.

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

Romer, P. 1990. "Endogenous Technical Change." *Journal of Political Economy* 98:S71–S102.

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

Schervish, M. 1995. *Theory of Statistics*. New York: Springer-Verlag.

Schultz, T. P. 1998. "Inequality in the Distribution of Personal Income in the World: How It Is Changing and Why." *Journal of Population Economics* 11:307–44.

———. 1999. "Health and Schooling Investments in Africa." *Journal of Economic Growth* 13:67–88.

Sims, C. 1980. "Macroeconomics and Reality." *Econometrica* 48:1–48.

Temple, J. 1998. "Robustness Tests of the Augmented Solow Growth Model." *Journal of Applied Econometrics* 13:361–75.

———. 2000. "Growth Regressions and What the Textbooks Don't Tell You." *Bulletin of Economic Research* 52:181–205.

Tierney, L., and J. Kadane. 1986. "Accurate Approximations for Posterior Moments and Marginal Densities." *Journal of the American Statistical Association* 81:82–6.

Wasserman, L. 2000. "Bayesian Model Selection and Model Averaging." *Journal of Mathematical Psychology* 44:97–102.

White, H. 1994. *Estimation, Inference and Specification Analysis*. Cambridge: Cambridge University Press.

Wolcott, S., and G. Clark. 1999. "Why Nations Fail: Managerial Decisions and Performance in Indian Cotton Textiles, 1890–1938." *Journal of Economic History* 59:397–423.