NBER WORKING PAPER SERIES

GROWTH ECONOMICS AND REALITY

William A. Brock
Steven N. Durlauf

Working Paper 8041
http://www.nber.org/papers/w8041

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2000

This paper was prepared for the World Bank Economic Review. We thank the National Science Foundation, John D. And Catherine T. MacArthur Foundation, Vilas Trust, and Romnes Trust for financial support. We thank François Bourguignon for initiating this research as well as Cullen Goenner, Andros Kourtellos, Arthur Minkin, Eldar Nigmatullin and seminar participants and Carnegie-Mellon and Pittsburgh for helpful comments. Chih Ming Tan has provided superb research assistance. Special thanks go to William Easterly and Ross Levine for sharing data and helping with replication of their results. The views expressed in this paper are those of the authors and not necessarily those of the National Bureau of Economic Research.

Growth Economics and Reality
William A. Brock and Steven N. Durlauf
NBER Working Paper No. 8041
December 2000
JEL No. O4, C1, C5

ABSTRACT

This paper questions current empirical practice in the study of growth. We argue that much of the modern empirical growth literature is based on assumptions concerning regressors, residuals, and parameters which are implausible both from the perspective of economic theory as well as from the perspective of the historical experiences of the countries under study. A number of these problems are argued to be forms of violations of an exchangeability assumption which underlies standard growth exercises. We show that relaxation of these implausible assumptions can be done 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 comprise draws from the same statistical model. We propose ways to account for both theory and heterogeneity uncertainty. Finally, using an explicit decision-theoretic framework, we describe how one can engage in policy-relevant empirical analysis.

William A. Brock
Department of Economics
University of Wisconsin
1180 Observatory Drive
Madison, WI 53706-1393
wbrock@ssc.wisc.edu

Steven N. Durlauf
Department of Economics
University of Wisconsin
1180 Observatory Drive
Madison, WI 53706-1393
and NBER
sdurlauf@ssc.wisc.edu

1. Introduction

> "There are more things in heaven and earth, Horatio,
> Than are dreamt of in your philosophy"
>
> Hamlet, act I, scene V

This paper has an ambitious objective – to outline a new perspective on empirical growth research which will both address some of the major criticisms to which 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 terms of empirical work, efforts to explain cross-country differences in growth behavior since World War II have become a predominant area of empirical work. Beyond the predominance of growth research within the academic community, the implications of this work for policymakers is immense. For example, there exist strong links between national growth performance and international poverty and inequality. Differences in per capita income across countries is substantially larger than those within countries, Schultz (1998) concludes that two-third's of (conventionally measured) inequality across individuals internationally is due to intercountry differences, so that efforts to reduce international inequality are naturally led to focus on cross-country growth differences. In turn, the academic community has used this new empirical work as the basis for very strong policy recommendations. A good example of this is Barro (1996) who concludes, based on a linear cross-country growth regression of the type that is so standard in this literature, 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..." (pg. 24)

At the same time, it seems clear that 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) 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." (pg. 68-69)

and Schultz (1999):

"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." (pg. 71)

In fact, it seems no exaggeration to say that the growth literature in economics is notable for the large gaps which continue to exist between theory and empirics. A recent (and itself critical) survey of the empirical literature, Durlauf and Quah (1999), concludes

"...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. Further, the new stylized facts of growth, as embodied in nonlinearities and distributional dynamics have yet to be integrated into full structural econometric analysis." (pg. 295)

The purpose of this paper is twofold. First, we attempt to identify some general methodological problems which we believe explain the widespread mistrust of growth regressions. While the factors we identify are certainly not exhaustive, they do in our judgment represent the most serious criticisms of conventional growth econometrics of which we are aware. These problems are serious enough to qualify and even 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, we believe, both from the perspective of economic theory and from the perspective of the historical experiences of the countries under study. We further show that these problems are interpretable in terms of model uncertainty and outline methods of reporting the level of uncertainty in our conclusions that more accurately reflect the "true" level of uncertainty consistent with the data.

Second, we argue that the use of empirical growth work for policy analysis should be reconceptualized along explicit decision-theoretic lines. Other authors, notably Levine and Renelt (1992), Fernandez, Ley, and Steel (1999), and Doppelhofer, Miller and Sala-i-Martin (2000) have each addressed some of the aspects of the model uncertainty which we believe underlies the mistrust of conventional growth regressions. At the same time, there has yet to be a treatment which is "policy-relevant" by which we mean an analysis of how model uncertainty impacts on the evaluation of alternative policies. Specifically, we are interested in understanding how one may employ cross-country growth data to compute predictive distributions for the consequences of policy outcomes, distributions which can then be combined with a policymaker's welfare function to assess alternative policy scenarios.

The methods we describe can, we believe, 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. As will be clear from our discussion, this paper only begins to scratch the surface of a policy-relevant growth econometrics. Our hope is that the ideas in the paper will facilitate new directions in growth research.

The title of this paper intentionally echoes Sims' classic (1980) critique of macroeconometric models. The growth literature does not suffer from the exact type of "incredible" assumptions (Sims (1980)) which were required to identify economic structure through 1960's-style simultaneous equations 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]

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

At the same time, our purpose is not to argue that statistical analyses of cross-country growth data are incapable of providing insights. In our judgment, there are several critical roles for regression and other forms of statistical analysis in the study of growth. One role is the identification of interesting data patterns, patterns which 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. Our focus, on the other hand, will explicitly be on the role of empirical work in formulating policy recommendations. In particular, the second goal of this paper is to explore how one can, by casting empirical analysis in an explicitly decision-theoretic framework, develop firmer insights into the growth process.

In our discussion, we will take an eclectic stance on how one should go about data analysis. As will be apparent, many of our ideas will be derived from the Bayesian statistics literature. Yet we think the basic arguments we make are relevant to frequentist analyses. Our view of data analysis is essentially pragmatic. Data analyses of the sort which are conventional in economics should, in our judgment, be thought of as evidence gathering exercises based on the goals of 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 mathematically 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 after the analysis is 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 by whether the level of uncertainty around a conclusion is low enough to be of policy relevance. Our position is really that one should evaluate various statistical procedures on the basis of whether they successfully answer the questions for which they are employed; we are unconcerned, at least in this paper, with abstract issues which distinguish frequentist and Bayesian approaches, for example. Because of this perspective, our criticisms of the empirical growth literature have no necessary implication for other empirical enterprises, although they are certainly potentially germane in similar cases.

2. 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). While recent work has extended growth analysis to consider panels (Evans (1998), Islam (1995), Lee, Pesaran, and Smith (1997)), the arguments we make concerning conventional empirical growth practice as well as our proposed alternative approach are equally relevant to that context as well, so we will focus on cross-sections.

A generic form for cross-country growth regressions is

$$g_i = y_{i,0}\beta + X_i\gamma + Z_i\pi + \epsilon_i \tag{1}$$

where g_i is real per capita growth in economy i over a given time period (typically measured as the change in per capita income between the beginning and end of the sample divided by the number of years which have elapsed), $y_{i,0}$ is the logarithm of per capita income of country i at the beginning of the time period under study, X_i is a vector of country-specific savings and population growth rate controls, Z_i is a vector of country-specific control variables, and ϵ_i is the error term. This decomposition of the possible control variables in a cross-country growth regression corresponds to different substantive aspects of modern growth theory. The variables $y_{i,0}$ and X_i represent those country-specific controls whose presence is predicted by the Solow (1957) growth model. The Solow model is very often treated as a baseline from which to build up more elaborate theories, hence these variables tend to be common across studies. Z_i captures any additional controls which a particular researcher believes are important and so generally differs across analyses. The singling out of $y_{i,0}$ as a determinant of growth is done in order to allow for assessment of the convergence hypothesis, i.e. the proposition that for two countries with identical levels of the X_i and Z_i controls, the one with lower initial output will on average grow faster than the other, so long as saving and population growth rates are equal.[2]

While this regression is typically applied to national aggregates, it is worth noting that 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 impact different regions, or population groups differently. Our conjecture is that such decompositions are important when evaluating growth policies with significant distributional consequences.

3. Econometric issues

In this section, we discuss three basic problems with the use of the basic equation (1). These problems each, at one level, occur because of violations of the assumptions necessary to estimate (1) using ordinary least squares 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 which motivate such analyses.

i. openendedness of theories

A fundamental problem with growth regressions is the determination of what variables to include in the analysis. This problem occurs because growth theories are openended. By openendness, 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 there exists a causal relationship between trade policy and growth. As a result, one finds that well over 90 different variables have been proposed as potential growth determinants (Durlauf and Quah (1999)), each one of which has some ex ante plausibility. As there are at best approximately 120 different countries available for analysis in cross-sections (the

[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 versus statistical meaning of convergence.

number may be far fewer due to missing observations on some covariates), it is far from obvious how to formulate firm inferences about any particular growth explanation.

This issue of openendness has not been directly dealt with in the literature. Rather, a number of researchers have proposed ways to deal with the robustness of variables in growth regressions. The basic idea of this approach is to identify a set of potential control variables for inclusion in (1) as elements of Z_i. Inclusion of a variable in the final choice of Z_i requires that its associated coefficient proves to be robust with respect to the inclusion of other variables. Levine and Renelt (1992) first introduced this idea, employing Leamer's (1978,1983) extreme bounds analysis. In extreme bounds analysis, a coefficient is robust if the sign of its OLS estimate stays constant across a set of regressions which represent different possible combinations of other variables. Sala-i-Martin (1997), arguing that extreme bounds analysis is likely to lead to the rejection of variables which actually do influence growth, proposed the standard that a coefficient is robust if it is statistically significant in 95% of the alternative regressions run.

These proposals for dealing with the plethora of growth theories are, in our judgment, 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 to not be robust if its associated variable is highly collinear with variables suggested by other candidate growth theories. This is obvious in the case of the Sala-i-Martin approach, since collinearity easily can produce statistical insignificance. This 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 along side it's corresponding regressor. Hence, these procedures will give sensible answers only if lack of collinearity is a "natural" property for a regressor which causally influences growth. Yet when one

[3]This remark is not meant to be patronizing. As we see it, 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 former we regard as fundamental and unassailable. The latter is a particular way of instantiating the deep idea 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.

thinks about theories of how various causal determinants of growth are themselves determined, it is clear that collinearity is a property one might certainly expect will hold for important causal determinants of growth.[4] This is easiest to see if one considers a block 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

$$g_i = d_i \gamma_d + \epsilon_i$$

$$d_i = c_i \pi_c + \eta_i \tag{2}$$

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.

Finally, we note that important recent papers by Doppelhofer, Miller and Sala-i-Martin (2000) and Fernandez, Ley and Steel (2000) have proposed to deal with regressor choice and hence at least indirectly with model openendedness through the use of Bayesian model averaging techniques. As this approach is very closely related to, and indeed helped to inspire our own views, we defer discussion of these papers until Section 6.

ii. 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 which describe growth are identical across countries. This seems to be a very implausible assumption. Does it really make sense to believe that the effect of a change in the level of a civil liberties index on growth in the United States is the same as that for Russia? While 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

[4]Leamer is in fact quite clear on this point. See Leamer (1978) pg. 172 for further discussion.

socioeconomic data set. After all, economic theory does not imply that individual units ought to be characterized by the same behavioral functions. That being said, any empirical analysis necessarily will require that there exists a set of interpretable statistical properties that are common across observations; it is a matter of judgment when homogeneity assumptions are or are not to be made. Our contention is that the assumption of parameter homogeneity seems particularly inappropriate when one is studying complex heterogeneous objects such as countries.

Evidence of parameter heterogeneity has been developed in different contexts, examples include Canova (1999), Desdoigts (1999), Durlauf and Johnson (1995), Durlauf, Kourtellos, and Minkin (2000), Kourtellos (2000) and Pritchett (2000). While these studies use very different statistical methods, each suggests that the assumption of a single linear statistical growth model which applies to all countries is incorrect.[5] To put it a different way, 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.[6]

***iii.* causality versus correlation**

A final source of skepticism with conventional growth empirics concerns a problem endemic to all structural inference in social science – the question of causality versus correlation. Many of the standard variables which are used to explain growth patterns – democracy, trade openness, rule of law, social capital, etc. are as much outcomes of socioeconomic relationships as growth itself. Hence there is an a priori case that the use of ordinary least squares estimates of the relationship between growth and such variables cannot be treated as structural anymore than coefficients which are produced by OLS regressions of price on quantity. 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

[5]Durlauf and Johnson (1995) provide substantial evidence of parameter heterogeneity between countries based on differences in initial literacy and initial income. Recently, conventional growth analyses have begun to pay some attention to differences between rich and poor countries; Barro (1996), for example allows the effects of democracy on growth to differ between rich and poor.

[6]See Draper (1997) for discussion of this idea.

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.

To be clear, 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, in our judgment, this trend towards the use of instrumental variables estimation has not satisfactorily addressed this problem. The reason is that the first criticism we have made of conventional growth econometrics – the failure to account properly for the openendness of growth theories – has important implications for validity of instrumental variables methods.

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

$$y_i = R_i\beta + \epsilon_i \tag{3}$$

the use of some set of instrumental variables I_i as instruments for R_i requires, of course, that each element of I_i is uncorrelated with ϵ_i. In the growth literature, this is not the condition that is actually employed to motivate instruments choice. Rather, instruments are typically chosen on the basis that they are exogenous, which operationally means that they are predetermined with respect to ϵ_i. Predetermined variables, however, are not necessarily valid instruments.

A good example of this pitfall may be found in Frankel and Romer (1996) which studies the relationship between trade and growth.[7] Frankel and Romer argue that since trade openness is clearly endogenous, it is necessary to instrument the trade openness variable in a cross-country regression if one is to consistently estimate the trade openness coefficient. In order to do this, they use a geographic variable, i.e. area, as an instrument and argue, perfectly sensibly, that area is predetermined with respect to growth. 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, etc. For example, larger countries may be more likely to be ethnically heterogeneous, leading to

[7] This discussion borrows from Durlauf (2000).

attendant social problems. Alternatively, larger countries may have higher per capita military expenditures, which means relatively greater shares of unproductive government investment and/or higher distortionary taxes. 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 one can defend such an assumption 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, the theory identifies instrumental variables, namely any variables which are observable at the time expectations are formed.

4. Exchangeability

Inferences from any statistical model can only be made, of course, conditional on various prior assumptions which translate the data under study into a particular mathematical structure. One way to evaluate the plausibility of inferences drawn from empirical growth regressions is by assessing the plausibility of the assumptions which are 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, it is typical to assume that the errors in a cross-section regression are jointly uncorrelated and orthogonal to the model's regressors.[8] Does one really wish to argue that there literally exist no omitted factors which induce correlation across the innovations in the growth regressions associated with the model?

[8] In the subsequent discussion, we will focus on OLS estimation of growth regressions. In the empirical growth literature, one can find examples of heteroskedasticity corrections to relax assumption 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.

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 and/or the regressor orthogonality assumption.

Estimation, inference, and the various uses of a statistical model such as (1) can still be conducted under weaker assumptions than uncorrelatedness and orthogonality. Econometrics has a long tradition of identifying minimal sets of conditions under which coefficient estimates and standard errors may be consistently estimated. Examples of this include the emphasis on orthogonality conditions between regressors and errors as the basis for ordinary least squares consistency (as opposed to 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 modelling of the series as a known ARMA process). Hence, any critique of cross-country growth analyses which is based upon the plausibility of particular statistical assumptions needs to argue that the violations of the assumptions actually 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 both be interpreted as examples of deviations of empirical growth models from a statistical "ideal" that allows for the sorts of inferences one wishes to make in growth contexts. The purpose for doing this is to establish a baseline for statistical growth models such that if a model does not meet this standard, then 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 per se, but rather helps define a strategy which we think empirical workers should follow in formulating growth models. As we see it, when a model does not meet this standard, a researcher must be prepared to argue that the violations of the standard do not invalidate the empirical claims he wishes to make. This standard is based on a concept in probability known as *exchangeability*.[9]

[9]Bernardo and Smith (1994) provide a complete introduction to exchangeability. Draper, Hodges, Mallows and Pregibon (1993) develop a detailed argument on the importance of exchangeability to statistical inference. Our analysis is very much indebted to their perspective.

i. **basic ideas**

A formal definition of exchangeability is the following.

Definition. Exchangeability

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

$$\mu(\eta_1 = a_1,...,\eta_K = a_K) =$$

$$\mu(\eta_{\rho(1)} = a_1,...,\eta_{\rho(K)} = a_K)^{10} \tag{4}$$

where $\rho(\cdot)$ is any operator which 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 an additional concept, which we call $\mathfrak{F}$ – conditional exchangeability.[11]

Definition. $\mathfrak{F}$ – conditional exchangeability

For a sequence of random variables η_i and collection of associated random vectors $\mathfrak{F}_i$, η_i is $\mathfrak{F}$ – conditionally exchangeable if, for every finite collection $\eta_1 ... \eta_K$ of elements of

[10]Throughout, $\mu(\cdot)$ will be used to denote probability measures.

[11]$\mathfrak{F}$ – conditional exchangeability is originally defined in Kallenberg (1982). Ivanoff and Weber (1996) provide additional discussion. The notion of $\mathfrak{F}$-conditional exchangeability is rarely employed in the statistics literature, and is not mentioned in standard textbooks such as Bernardo and Smith (1994). We believe this is so because exchangeability analyses in the statistics literature generally focus on whether the units under study are exchangeable, rather than the units conditional on certain characteristics as is the more natural notion in economic contexts.

the sequence,

$$\mu(\eta_1 = a_1,...,\eta_K = a_K \mid \underset{\sim}{\mathfrak{F}}) =$$

$$\mu(\eta_{\rho(1)} = a_1,...,\eta_{\rho(K)} = a_K \mid \underset{\sim}{\mathfrak{F}}) \tag{5}$$

where $\rho(\cdot)$ is any operator which permutes the K indices and $\underset{\sim}{\mathfrak{F}} = \{\mathfrak{F}_i\}$.

Notice that when $\forall i$, $\mathfrak{F}_i = \phi$, the empty set, $\mathfrak{F}-$conditional exchangeability reduces to exchangeability.

Associated with exchangeability and $\mathfrak{F}-$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 $\eta_1...\eta_K$ of elements of the sequence,

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

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

The key difference between exchangeability and partial exchangeability is the conditioning on common values of some random vectors Y_i's associated with the η_i's in the latter case. If Y_i is a discrete variable, then partial exchangeability implies that a sequence may be decomposed into a finite or countable number of exchangeable subsequences.

How does exchangeability relate to the assumptions which underlie cross-country growth regressions? These models typically assume that once the included growth variables in the model are realized, there exists no basis for distinguishing the probabilities of various permutations of residual components in country-level growth rates, i.e. these residuals are $\mathcal{F}$ – conditionally exchangeable, where $\mathcal{F}$ is the modeller's information set. Notice that knowledge of variables outside those included in a regression or parameter heterogeneity represent components of this information set.

Even though $\mathcal{F}$ – conditional exchangeability of model errors constitutes a stronger assumption than is needed for many of the interpretations of ordinary least squares, this particular exchangeability condition is nevertheless very useful as a benchmark in the construction and assessment of statistical models. We make this claim for two reasons. First, this exchangeability concept helps to organize discussions of the plausibility of the invariance of conditional moments that lie at the heart of policy relevant predictive exercises. Draper (1987) 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." (pg. 458)

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

Second, the conditions under which various types of exchangeability do or do not hold for growth rates or model residuals can be linked to one's substantive understanding of the growth process in ways that alternative sets of (purely statistical) assumptions on errors cannot. In turn, once exchangeability is believed to be violated, one can naturally link the reasons why 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.

These considerations are fundamental to questions of model formulation. Further, exchangeability is separately important in terms 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:[12] 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. 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.[13]

de Finetti's Representation Theorem.

If η_i is an infinite exchangeable sequence with associated probability measure P, then there exists a probability measure Q over $\mathcal{F}$, 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

$$F(\eta_{i-j}\ldots\eta_i\ldots\eta_{i+k}) = \int \prod_{r=-j}^{k} F(\eta_{i+r})dQ(F). \tag{7}$$

What is important about de Finetti's Representation Theorem in terms of empirical practice is that it creates a link between a researcher's prior beliefs about the nature of the data under analysis, specifically the properties of regression errors, which permits one to interpret ordinary least squares estimates and the test statistics in the usual way.[14]

***ii.* exchangeability and growth**

Various forms of exchangeability appear, in our reading of the empirical growth literature, to implicitly underlie many of the regression specifications one encounters. An implicit ($\mathcal{F}$ – 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)

[12]See Bernardo and Smith (1994) chapters 4 and 6 for a very insightful discussion of the nature and implications of the theorem and Aldous (1983) for a comprehensive mathematical development of various forms of the theorem.

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

by allowing the trajectory's state variables to differ across countries and the appending of an error term. Such an assumption of exchangeability has substantive implications for how one thinks about the relationship between a given observation and others in a data set. Suppose that one is considering the effect of a change in trade openness on a country, e.g. Tanzania, in sub-Saharan Africa. How does one employ estimates of the effects of trade on growth in other countries to make this assessment? The answer to this question 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 modelling assumptions which are standard in conventional growth exercises are conceptually related to the assumptions that the model errors ϵ_i are 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 $\mathcal{F}-$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

[14] One needs to be cautious in the use of di Finetti's theorem to calculate the distributions of regression estimators. For linear regression models of the form (1) with normally distributed errors and nonstochastic regressors, Arnold (1979) 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.*" (pg. 194)

Further, if the errors are nonnormal, then 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 $\mathcal{F}-$conditional exchangeability, although as far as we know no such results have been established.

that the errors in a growth regression are exchangeable (or are exchangeable after a heteroskedasticity correction.)

Our general claim is that exchangeability, in particular, $\mathcal{F}-$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 eq. (1), in which a small number of regressors are assumed to explain cross-country growth patterns.[15]) In order for exchangeability to hold, it would be necessary for a given regression and modeller's information set that the likelihood of a positive error for a given country, say Japan, to be the same as that of any other country in the sample. In turn, for this to be true, it would have to be the case that no prior information exists about the countries under study which renders 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. Hence our use of the term "related" in the three points we enumerate 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. When a researcher estimates a model for which exchangeability is violated, it is incumbent on the researcher to explain how the estimation and interpretation of the model is consistent with these violations.

The failure of exchangeability to hold in cross-country contexts may seem to be an esoteric objection, but in fact it speaks to a fundamental issue in the interpretation of cross-country growth regressions, namely, the leap from the identification of statistical patterns to causal inference. Suppose 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 non-negligible variance), or conclude that there was something about those countries which

[15]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.

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), based on prior knowledge about the politics and cultures of these countries, developed their analysis on the basis that the Solow errors were not exchangeable, i.e. that there was something about the sub-Saharan African countries which should have been incorporated into the Solow model.

An important question for the assessment of growth regressions is whether there can ever exist a point where our knowledge of the history and culture of the countries involved in cross-country regressions is embedded in the regressions to such an extent that the exchangeability requirement is met. Notice, this is a deeper issue than asking whether the covariance matrix of residuals for a model is proportional to the identity matrix (or even whether it is diagonal). In the growth context, the failure of exchangeability means that a modeller has reasons to question whether the countries in a sample obey the same growth process.

Does the requirement of exchangeability imply the impossibility of structural inference whenever observational data are being studied? In our view, 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 very plausible assumption for statistical models based on the use of individual level data sets such as the PSID or NLSY 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 micro data sets of this type and macro data sets of the type used in growth analysis is the latter observations pertain to large heterogeneous aggregates for which a great deal of information is known which implies that exchangeability does not hold. In addition, the large size of individual level data sets such as the PSID or NLSY means that the range of possible control variables is much greater than is the case for growth. By this we mean something deeper than "the more data points the more regressors which may be included." Rather, we would 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.

To summarize, in our view conventional growth econometrics has failed to consider the ways in which appropriate exchangeability concepts may or may not hold for the specific models which are analyzed. This failure in turn renders these studies difficult if not impossible to interpret since one must know whether any exchangeability violations which are present invalidate the statistical exercise being conducted. We therefore concur with Draper, Hodges, Mallows and Pregibon (1987), 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." (pg. 1)

iii. exchangeability and causality

While exchangeability is, we believe, a useful benchmark for understanding some of the major sources of skepticism in growth regressions, it should be emphasized that 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. The reason for this is simple: exchangeability relates to the relationship between joint probabilities of outcomes and not to anything concerning joint probabilities for different variables under alternative probability distributions for other variables. 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.

Put differently, causality is a different sort of question from 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) 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." (pg. 89)

and

"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." (pg. 91)

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

5. A digression on noneconometric evidence

Regression analyses of the type which are 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. We do this as these sources are often the basis on which one can question exchangeability in a particular context.

The economic history literature is replete with studies which are of enormous importance in adjudicating different growth explanations, yet this literature really only receives lip service in growth literature.[16] An exemplar of the sort of historical study which speaks to growth debates is Clark (1987). This paper explores the sources of productivity differences between cotton textile workers in New England versus workers in other countries in 1910. These differences were immense – a typical New England worker was about 6 times as productive as his counterpart in India or China and over twice as productive as his counterpart in Germany. Clark painstakingly shows that these differences cannot be attributed to differences in technology, education, or management.[17] Rather, they seem to reflect cultural differences in work and effort norms. This type of study has 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 than is done currently.

Further, we believe that historical and qualitative studies play a crucial role in the development of credible statistical analyses. One reason for this is that such studies provide information on the plausibility of identifying assumptions which are made to establish causality. Further, our discussion on exchangeability and growth analysis may be interpreted as arguing it is necessary that a researcher be able to do one of two things in order to argue that a regression provides structural information. A researcher can make a plausible argument that, given the many plausible growth theories and plausible heterogeneity in the way different causal growth factors impact on different countries, the errors in particular growth regression are nevertheless exchangeable. Or, a researcher must make the argument that the violations of exchangeability in the regression occur in ways which do not affect the interpretation of the coefficients and standard errors from those which are employed. To some extent, exchangeability judgments must be made prior to a statistical exercise, as noted by Draper et al (1987) in the quotation which appears above. Where does information of this type come from? In our judgment, it is qualitative and historical work which provides such information. Hence, the detailed study of individual countries which is a hallmark of work by the World Bank, for example, plays an invaluable role in allowing for credible statistical analysis.

[16]There are of course a number of notable exceptions to this remark, such as Easterly and Levine (1997) and Prescott (1998).

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

6. Towards a policy-relevant growth econometrics

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

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

which implies

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

and

3) Proposed solutions to these problems have typically been ad hoc.

These problems need to be explicitly dealt with in order for econometric inferences on growth to be credible. More specifically, we advocate an explicit decision-theoretic basis for assessing growth data. As we see it, when one attempts to analyze a relationship such as that between democracy and growth, one is ultimately interested in policy-based predictions, i.e. the construction of predictive densities to assess the effects of changes in policy. The decision-theoretic framework we employ is explicit about various forms of model uncertainty which are associated with possible violations of exchangeability, as we have discussed in the previous section.

From the decision-theoretic perspective, one moves away from a specific concern with a particular hypotheses to an evaluation of the implications of a given set of data for a particular course of action on the part of the researcher. As argued by Kadane and Dickey (1980),

"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 turns 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 inappropriate 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" (pg. 247)

When one substitutes "policymaker" for "scientist" in this quotation, it becomes clear why policy-relevant growth econometrics needs to explicitly integrate policy objectives and empirical practice. Our particular approach is well summarized by Kass and Raftery (1994):

"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." (pg. 784)

In other words, we argue that policy-relevant econometrics needs to explicitly identify the objectives of the policymaker, and then calculate the expected consequences of a policy change.[18] Within this calculation, explicit accounting needs to be taken of model uncertainty. This approach is related to Doppelhofer, Miller and Sala-i-Martin (2000) and Fernandez, Ley and Steel (1999); explicit links will be described below.

It is important to recognize that there are no panaceas for the interpretation problems we have described for growth regressions. While our formulation of model uncertainty can, we believe reduce the dependence of empirical growth studies on untenable exchangeability or other assumptions, these procedures will still rely on untestable and possibly controversial assumptions. It may be impossible, for example, to place every possible growth theory in a common statistical analysis, so critiques based on theory openendness 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 et al (1993) refer to as "the role of leaps of faith" (pg. 16) in constructing statistical models. What we wish to do is reduce the number and magnitude of such leaps.

[18]See Chamberlain (2000) for an extensive development of econometric machinery for facilitating decisionmaking under uncertainty.

i. **modelling model uncertainty**

We will assume that the structural growth process for country i obeys a linear structure which applies to all countries j which are members of a class $\mathfrak{J}(i)$. Suppose that this model is described by a set of regressors S which we partition into a subset X and a scalar z. Our analysis will focus on analyses based on β_z, the coefficient which determines the role of z_i on country i's growth. Hence we work with models of the form

$$g_j = S_j\zeta + \epsilon_j = X_j\pi + z_j\beta_z + \epsilon_j,\ j \in \mathfrak{J}(i). \tag{8}$$

When a given model represents the "true" or correct specification of the growth process for country i, then the sequence of residuals ϵ_j will be $\mathfrak{F}-$exchangeable where $\mathfrak{F}$ constitutes all potential growth regressors (which will include both S and any variables which capture parameter heterogeneity as described below). 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 $(\mathfrak{J}(i))$, which together render the associated model errors $\mathfrak{F}-$conditionally exchangeable, parallels our earlier discussion of the first two sources of criticisms of growth regressions. It is skepticism over the claim that a particular model is correctly specified which renders many of the empirical claims in the growth literature not credible.

In order to analyze a model of this type, it is necessary to add some statistical assumptions. We assume that ϵ_j's are jointly normal; the marginal distribution of the typical element is $N(0,\sigma_\epsilon^2)$. The variance σ_ϵ^2 is treated as known. Further, we will assume that conditional on a model, the various regressors in (8) are nonstochastic; the regressors are also assumed to be linearly independent. These assumptions will allow us to describe various posterior distributions in a simple fashion; the statistical literature has many ways of relaxing both assumptions. We denote the ordinary least squares estimate (as well as classical maximum likelihood estimate[19]) of ζ for a given model as $\widehat{\zeta}$. Finally we assume a noninformative prior on the regression coefficients,[20]

$$\mu(\zeta) \propto c. \tag{9}$$

Letting S denote the data matrix of regressors in (8), it is well known (eg. Box and Tiao (1973) pg. 115), that the posterior density $\mu(\zeta \mid D,M)$ is, for known σ_ϵ^2, multivariate $N(\hat{\zeta},(S'S)^{-1}\sigma_\epsilon^2)$; the posterior density of β_z can of course be calculated from this vector distribution. One can think of the standard approach to statistical analysis in the growth literature as using a single model M and given data set D to analyze model parameters. 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 means calculating the posterior density $\mu(\beta_z \mid D,M)$. Under the assumptions we have placed on the errors ϵ_j and using the noninformative prior (7), the posterior mean and variance of the parameters, which are the natural objects to compute under the Bayesian approach, will equal the corresponding OLS parameter and variance estimates. This equivalence will allow us to keep the interpretation of model uncertainty clear from both Bayesian and frequentist perspectives.

Now suppose that there exists a universe of models, $M_m \in \mathcal{M}$, which are possible candidates for the "true" growth model which generated the data under study. This universe is generated from two types of uncertainty. First, there is *theory uncertainty.* In particular, we assume that there is a set $\mathfrak{S}$ of possible regressors to include in a growth regression which 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 (8). Second, there is *heterogeneity uncertainty.* By this, we mean that there is uncertainty as to which countries comprise $\mathfrak{J}(i)$, i.e. are partially

[19] As was pointed out in footnote 13 above, Arnold (1979) shows that for regressions of the form (1), with normal errors and nonstochastic regressors, the best linear unbiased estimator is the same as the ordinary least squares estimator regardless of whether the errors are i.i.d. or exchangeable. Hence minimum variance estimation is achieved for the regression parameters for (1).

[20] We use this noninformative prior exclusively for expositional purposes. This prior has the disadvantage that it is "improper," i.e. it integrates to ∞ rather than 1. For this reason, the calculation of certain quantities such as Bayes factors is precluded. See Raiffa and Schlaifer (1961) for examples of proper priors such as the normal/Gamma which may be used in analyzing the posterior distribution of linear regression coefficients. In our empirical work, we use an approximation due to Raftery (1995) in which the prior coefficient distribution can be ignored.

exchangeable with country i.[21] In the presence of these types of uncertainty, one will not be interested in $\mu(\beta_z \mid D, M_m)$ for a particular M_m but rather in $\mu(\beta_z \mid D)$; the exception, of course, is when the correct model M_m and the set of countries which are partially exchangeable with i are both known with probability 1 to the modeller.

It is possible to integrate theory and some forms of heterogeneity uncertainty into a common variable selection framework. To see this, for a given regressor set S_k, suppose that one believes that the countries under study may be divided into two subsets with associated subscripts $\mathcal{I}_1$ and $\mathcal{I}_2$ 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

$$g_j = X_j\pi_1 + z_j\beta_{z,1} + \epsilon_j \text{ if } j \in \mathcal{I}_1 \tag{10}$$

and

$$g_j = X_j\pi_2 + z_j\beta_{z,2} + \epsilon_j \text{ if } j \in \mathcal{I}_2 \tag{11}$$

This last equation can be rewritten as

$$g_j = X_j\pi_1 + z_j\beta_{z,1} + X_j(\pi_2 - \pi_1) + z_j(\beta_{z,2} - \beta_{z,1}) + \epsilon_j \text{ if } j \in \mathcal{I}_2. \tag{12}$$

Therefore, one can combine the two equations into a single growth regression of the form

$$g_j = X_j\pi_1 + z_j\beta_{z,1} + X_j\delta_{j,\mathcal{I}_2}\pi_{\mathcal{I}_2} + z_j\delta_{j,\mathcal{I}_2}\beta_{z,\mathcal{I}_2} + \epsilon_j,\ j \in \mathcal{I}_1 \cup \mathcal{I}_2 \tag{13}$$

where $\delta_{j,\mathcal{I}_2} = 1$ if $j \in \mathcal{I}_2$, 0 otherwise, and one will have obtained a regression in which $X_j\delta_{j,\mathcal{I}_2}$ and $z_j\delta_{j,\mathcal{I}_2}$ are now defined over all j, they are simply additional regressors. Notice that it is straightforward to generalize this procedure to multiple groups of partially exchangeable countries. To be clear, this procedure is not completely general in that it restricts the sort of possible parameter heterogeneity which is allowed; for

[21] These two types of uncertainty are not independent.

example, each country is not allowed a separate set of coefficients. In order to allow for this more general type of heterogeneity, it would be necessary to move to an alternative structure, such as a hierarchical linear model (see Schervish (1995), chapter 8); we plan to pursue this in subsequent work.

In the presence of model uncertainty, the calculation of $\mu(\beta_z \mid D)$ requires that one integrates out the dependence of the probability measure $\mu(\beta_z \mid D, M_m)$ on the model M_m. By Bayes' rule, one can compute the posterior distribution of a given coefficient conditional only on the observed data,

$$\mu(\beta_z \mid D) = \sum_m \mu(\beta_z \mid D, M_m)\mu(M_m \mid D) \tag{14}$$

where $\mu(M_m \mid D)$ denotes the posterior probability (given the data) of model M_m. Further, since

$$\mu(M_m \mid D) \propto \mu(D \mid M_m)\mu(M_m) \tag{15}$$

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 expression may be rewritten as

$$\mu(\beta_z \mid D) \propto \sum_m \mu(\beta_z \mid D, M_m)\mu(D \mid M_m)\mu(M_m) \tag{16}$$

When the prior probabilities $\mu(M_m)$ are equal, then

$$\mu(\beta_z \mid D) \propto \sum_m \mu(\beta_z \mid D, M_m)\mu(D \mid M_m) \tag{17}$$

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) pg. 118 gives the following derivations of the conditional mean and variance of β_z given the data D:

$$E(\beta_z \mid D) = \sum_m \mu(M_m \mid D) E(\beta_z \mid D, M_m) \tag{18}$$

and

$$var(\beta_z \mid D) = E(\beta_z^2 \mid D) - (E(\beta_z \mid D))^2 =$$

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

$$\sum_m \mu(M_m \mid D) var(\beta_z \mid M_m, D) + \sum_m \mu(M_m \mid D)(E(\beta_z \mid D, M_m) - E(\beta_z \mid D))^2 \tag{19}$$

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 (19) and the variance of the estimates across models (the second term in (19)).[22]

Equation (14) and the related expressions are all examples of Bayesian model averages. The methodology surrounding Bayesian model averaging is specifically developed for linear models in Raftery, Madigan, and Hoeting (1997).[23] Doppelhofer, Miller and Sala-i-Martin (2000), focusing on theory uncertainty only, compute a number of measures of variable robustness based on applying 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, with a focus on the explicit computation of posterior coefficient distributions. Our own development differs from these important papers in two main respects. First, we treat heterogeneity uncertainty as well as theory uncertainty as part of overall model uncertainty. Draper, Hodges, Mallows and Pregibon (1993) provides a general overview of the importance of accounting for heterogeneity uncertainty in constructing credible empirical exercises. Second, we develop an explicit decision-theoretic approach to interpreting growth regressions.

[22] See Stewart (1984) for the development of some significance tests based on the posterior distribution of β_z.

[23] The survey by Hoeting, Madigan, Raftery, and Volinsky (1999) provides a nice introduction to model averaging techniques. See also Wasserman (1997).

ii. **policy assessment: basic ideas**

The basic posterior coefficient distribution (14) and associated first and second moments (18) and (19) represent data summaries and as such have no implications for either inference or policy assessment. The goal of a policy analysis is not the construction of such summaries but rather the assessment of the consequences of changes in the 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 done in extreme bounds analysis or the comparison of models using Bayes factors[24], requires the imposition of new evaluative criteria in order to produce explicit decision rules.

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, we wish to show 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 that the ignoring of regressors is shown to flow 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[25] in some scalar variable which is under the control of a policymaker and which is believed to have some effect on growth. Therefore, the decisionmaker's set of actions $\mathcal{A}$ is $\{0, dz_i\}$. This decision rule is based on a vector of observable data D. This means that a decisionmaker chooses a rule $\phi(\cdot)$ which maps D to $\mathcal{A}$ so that

$$\phi(D) = dz_i \text{ if } D \in D_1$$

$$\phi(D) = 0 \text{ otherwise} \tag{20}$$

[24]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.

[25]Without loss of generality, we will generally assume that $dz_i > 0$.

D_1 is therefore the acceptance region for the policy change. Finally, we reiterate that we assume that the "correct" linear growth model is a causal relationship.

Since we are restricting 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. From this perspective, 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 which 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 let $g_i + \beta_z dz_i$ denote the growth rate with the change. Finally, let $V(g_i)$ denote the utility value of the growth rate to the policy maker.

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

$$E(V(g_i + \beta_z dz_i) \mid D) - E(V(g_i) \mid D) \tag{21}$$

Calculations of the expected utility differential (19) implicitly contain all information that is relevant to a policy assessment. Hence, from the perspective of policy evaluation the various rules which have been proposed for the evaluation of regressor robustness should be an implication of this calculation. Notice that this calculation requires explicitly accounting for model uncertainty, since the conditioning is always down with respect to the data alone.

iii. policy assessment under alternative payoff functions

a. risk neutrality

Suppose that V is linear and increasing, i.e.

$$V(g_i) = \alpha_0 + \alpha_1 g_i,\ \alpha_1 > 0 \tag{22}$$

For a risk neutral 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, i.e.

$$\sum_m \mu(M_m \mid D) E(\beta_z \mid D, M_m) > 0 \tag{23}$$

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

$$\sum_m \mu(D \mid M_m) E(\beta_z \mid D, M_m) > 0 \tag{24}$$

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

b. mean/variance utility over possible changes

Suppose that a policy maker prefers higher expected growth, but dislikes variance in growth. For expositional simplicity, further assume that the payoff function of the policymaker only depends on the growth effect of the policy change. One way of expressing this is to assume that

$$E(V(g_i + \beta_z dz_i) \mid D) - E(V(g_i) \mid D) =$$

$$\alpha_0 E(\beta_z dz_i \mid D) + \alpha_1 var(\beta_z dz_i \mid D)^{1/2},\ \alpha_0 > 0,\ \alpha_1 < 0 \tag{25}$$

When $|\alpha_0/\alpha_1| = \frac{1}{2}$, then this payoff specification implies that the policy maker will only act if the t-statistic (i.e. the expected value of β divided by its standard deviation) is greater than 2. Hence, this specification corresponds to the standard econometric practice of ignoring regressors whose associated t-statistics are less than 2.

Notice that 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 only be justified under very special cases. First, it is necessary to assume that the form of risk aversion of the policy maker applies to the standard deviation rather than 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 payoff function is

$$E(V(g_i+\beta_z dz_i) \mid D) - E(V(g_i) \mid D) =$$

$$\alpha_0 E(\beta_z dz_i \mid D) + \alpha_1 var(\beta_z dz_i \mid D),\ \alpha_0 > 0,\ \alpha_1 < 0 \tag{26}$$

with $|\alpha_0/\alpha_1| = \frac{1}{2}$, then there will be a threshold level T such that for all $0 < dz_i \leq T$ a policy change increases the policymaker's utility.[26] Hence the rule of ignoring regressors with t-statistics less than 2 presupposes a very specific assumption on how risk affects the policymaker's utility. Second, if (25) 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, and alternatively may decline to act when the coefficient is statistically significant. These possibilities can be generated through appropriate choices of $|\alpha_0/\alpha_1|$.

c. Knightian uncertainty and maximin preferences

In the examples we have studied thus far, we have allowed all uncertainty concerning 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 that is in the tradition of Knightian uncertainty, assumes that an additional layer of uncertainty exists in the environment under study which may be interpreted as a distinct type of risk, as will be seen below.

Let $a=1$ when dz_i is undertaken, 0 otherwise. As before, let $\mathcal{M}$ denote the universe of possible growth models. One can define a risk sensitive payoff function for the policymaker as $W(a)$ where

$$W(a) =$$

[26] This is an example of the famous result of Pratt that one will always accept a small amount of a fair bet.

$$(1-e)E(V(g_i + a\beta_z dz_i) \mid D) + e(inf_{M_m \in \mathcal{M}} E(V(g_i + a\beta_z dz_i) \mid D, M_m)) \qquad (27)$$

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 classical Ellsberg paradoxes. For example, if experimental subjects are given a choice between 1) receiving 1 dollar if they draw a red ball at random from an urn which they know contains 50 red balls and 50 black balls or 2) receiving 1 dollar if they draw a red ball when the only information available is that it contains 100 red and black balls, the subjects typically choose the first, "unambiguous" urn (Camerer (1995) pg. 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.[27]

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 which exhibit such aversion. One popular representation (cf. Epstein and Wang) is to replace expected utility calculations of the form $\int u(\omega)dP(\omega)$ with $inf_{P \in \mathcal{P}} \int u(\omega)dP(\omega)$, where $\mathcal{P}$ is a space of possible probability measures. When this space contains a single element, this latter calculation reduces to the standard expected utility formulation. A variant of this formulation is to assume that the set $\mathcal{P}$ consists of a set of mixture distributions $(1-e)P_0 + eP_1$, where P_0 is a baseline probability measure that a decisionmaker 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 we have described, one can replace P by M_m and $\mathcal{P}$ by $\mathcal{M}$ and obtain the second part of (27). In using this particular specification, we do not claim that it is the only sensible way to model aversion ambiguity by policymakers. Rather, we introduce it so as to permit an illustration of 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

[27] We plan to address this for general policy choices in future work.

and increasing, that (23) holds. In this case $W(1) > W(0)$ if

$$(1-e)E(\beta_z \mid D) + e(inf_{M_m \in \mathcal{M}} E(\beta_z \mid D, M_m)) > 0 \tag{28}$$

When $e = 1$, the policy action will only be taken when $E(\beta_z \mid D, M_m) > 0$ for all $M_m \in \mathcal{M}$. Recall that under our assumption that the errors in the growth regressions are normal with known variance, the posterior expectation $E(\beta_z \mid D, M_m)$ equals $\widehat{\beta}_{z,m}$, the OLS coefficient associated with the regressor z for model M_m. Hence, when $e = 1$, one will only choose to implement dz_i if the OLS coefficient estimate of β_z is positive for every one of the models in $\mathcal{M}$.

Alternatively, assume that the policymaker is risk averse in the sense that (25) describes his payoff function. In this case, $W(1) - W(0)$ equals

$$(1-e)(\alpha_0 E(\beta_z dz_i \mid D) + \alpha_1 var(\beta_z dz_i \mid D)^{1/2}) +$$

$$e(inf_{M_m \in \mathcal{M}}(\alpha_0 E(\beta_z dz_i \mid D, M_m) + \alpha_1 var(\beta_z dz_i \mid D, M_m)^{1/2})) \tag{29}$$

In this case, the policy will be implemented if

$$(1-e)(\alpha_0 E(\beta_z \mid D) + \alpha_1 var(\beta_z \mid D)^{1/2}) +$$

$$e(inf_{M_m \in \mathcal{M}}(\alpha_0 E(\beta_z \mid D, M_m) + \alpha_1 var(\beta_z \mid D, M_m)^{1/2})) > 0 \tag{30}$$

Notice in this case, if $e = 1$ and $\mid \alpha_0/\alpha_1 \mid = 1/2$, then one would not act unless the regression coefficient $\widehat{\beta}_{z,m}$ is positive and statistically significant (again, in the sense that the t-statistic associated with its posterior density is at least 2) for each model in $\mathcal{M}$.

The policy rules which hold for $e = 1$ are closely related to the recommendations made by Leamer for assessment of coefficient fragility through extreme bounds analysis (see Leamer and Leonard (1978), Leamer (1978), and especially Leamer (1983)). In extreme bounds analysis, recall that when a regressor "flips signs" across specifications, this is argued to imply the regressor is fragile. From the perspective of policy recommendations, we interpret concern over fragility to mean that there should not be any policy change 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.

Finally, we would note that this derivation of extreme value analysis appears to answer a number of the objections raised to it by Granger and Uhlig (1985) and McAleer, Pagan and Volker (1985). Specifically, both of these papers argue that extreme bounds analysis can lead to spurious rejections of regressors due to changes in sign induced by regressions which are, by standard tests, misspecified. We agree that this is a problem with extreme bounds analysis as conventionally practiced. However, in our view, issues of model selection are appropriately subsumed in the computation of the posterior weights $\mu(M_m \mid D)$ and state (i.e. model) dependent utilities. As pointed out by Berger and Pericchi (1999), when one is engaged in Bayesian model selection between two models, one is in essence comparing the Bayes factors for the two,[28] which, if the prior odds on the models are equal, will equal the ratio of their respective likelihoods. The claim that a certain model is misspecified is equivalent to the claim that possesses a relatively lower likelihood relative to some other model that is correctly specified. In our context, misspecified models will receive lower posterior weights in the computation of the utility change defined by (30) since $\mu(M_m \mid D) \propto \mu(D|M_m)\mu(M_m)$. In this sense, we avoid the criticism of Bayes' factors in Dickey and Kadane (1980) that the weights do not account for the purpose of the empirical exercise – a criticism that applies to the criticisms of extreme bounds analysis we have described.

d. alternative utility functions

In the previous section, we assumed that the policymaker only cares about the level or change in growth induced by a policy change. It is of course possible to imagine other plausible payoff functions for a policymaker. One possibility is to

[28]See Camerer (1995) for additional examples of ambiguity aversion as well as a survey of the experimental economics literature's implication for utility theory.

assume that a policymaker evaluates a payoff based on changes in the expected value of growth within regime, i.e. that there exists a function ψ such that

$$\psi(E(g_i+\beta_z dz_i \mid D,M_m)-E(g_i \mid D,M_m)) \tag{31}$$

measures the payoff for a policy change conditional on a particular growth model. Linearity of the growth process of course implies that

$$E(g_i+\beta_z dz_i \mid D,M_m)-E(g_i \mid D,M_m)=E(\beta_z dz_i \mid D,M_m). \tag{32}$$

so the expected payoff from a fixed change dz_i, once one accounts for model uncertainty, is

$$\begin{aligned} E(\psi(E(\beta_z dz_i \mid D,M_m)) \mid D) &= \sum_m \mu(M_m \mid D)\psi(E(\beta_z dz_i \mid D,M_m)) \\ &= \sum_m \mu(M_m \mid D)\psi(\widehat{\beta}_{z,m} dz_i) \end{aligned} \tag{33}$$

recalling that $E(\beta_z \mid D,M_m)=\widehat{\beta}_{z,m}$. When $\psi(\cdot)$ is linear, this reduces to the risk neutral case discussed earlier. On the other hand, alternative payoff 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, then one will have an implied decision rule which says that a single sign change in the coefficient estimate $\widehat{\beta}_{z,m}$ as one moves across models is sufficient to imply that the policymaker should not act. This type of payoff function induces behavior which mimics that found under Knightian uncertainty.

At first glance, this might appear to be an unreasonable utility function for a policymaker. We think that this reaction is at least partially incorrect. Suppose each state of the world is indexed by the growth process that is "true" under it. Then, the utility of the policymaker will depend both on the growth rate which is expected to prevail as well as the state of the world under which it transpires. For example, suppose there is a model of the world in which the expected effect of democracy on growth is negative. Such a model of the world could be one whose features imply that a policymaker is particularly wary of reducing growth through 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

$$U(E(\beta_z \mid D,M_m)dz_i,M_m) = U(\widehat{\beta}_{z,m}dz_i,M_m) \tag{34}$$

so that the posterior expected utility of the policy change is

$$E(U(\widehat{\beta}_{z,m}dz_i,M_m) \mid D) = \sum_m \mu(M_m \mid D)U(\widehat{\beta}_{z,m}dz_i,M_m). \tag{35}$$

Manipulating $U(\cdot,\cdot)$ one can produce a result which is consistent with refusing to act whenever the posterior mean $\widehat{\beta}_{z,m}$ 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, $\widehat{\beta}_{z,m}$ is positive.

7. An empirical example[29]

In this section, we reconsider an important growth study, Easterly and Levine (1997) which examines the role of ethnic conflict in growth. We chose to reexamine this study because 1) it is generally regarded as quite important in the growth literature, 2) it has important implications for policy and the sorts of advice and advocacy an international organization would engage in, and 3) the authors of this study have done an extremely admirable job of making their data and programs publicly available.

Easterly and Levine's analysis is designed to understand why in standard cross

[29] It is a common Bayesian criticism of frequentist methods that the use of conventional confidence intervals is ad hoc.

Table 1. Data

All data are the same as used in Easterly and Levine (1997). The variables employed are:

GYP: Growth Rate of real per capita GDP.

AFRICA: Dummy variable for sub-Saharan African countries, as defined by the World Bank. These countries are Angola, Benin, Burundi, Botswana, Burkina Faso, Cameroon, Cape Verde, Central African Republic, Chad, Comoro Islands, Congo, Côte d'Ivoire, Djibouti, Equatorial Guinea, Ethiopia, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mauritius, Mozambique, Namibia, Niger, Nigeria, Rwanda, Sao Tome and Principe, Senegal, Seychelles, Sierre Leone, Somalia, South Africa, Sudan, Swaziland, Tanzania, Togo, Uganda, Zaire, Zambia, and Zimbabwe.

ASSASS: Number of assassinations per 1000 population.

BLCK: Black market premium, defined as log of 1 + decade average of black market premium.

DUM60: Dummy Variable for 1960's.

DUM70: Dummy Variable for 1970's.

DUM60: Dummy Variable for 1980's.

ELF60: Ethnic diversity measure; Index of ethnolinguistic fractionalization, 1960. This variable measures the probability that two randomly selected individuals from a given country will not belong to the same ethnolinguistic group.

LATINCA: Dummy variable for countries in Latin America and the Caribbean

LLY: Financial depth, measured as 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 measure at the start of each decade.

LRGDPSQ: Square of LRGDP.

LSCHOOL: Log of 1+ average years of school attainment, quinquennial values (1960-1965, 1970-1975, 1980-1985).

SURP: Fiscal Surplus/GDP: Decade average of ratio of central government surplus to GDP, both in local currency, local prices.

Table 2 : OLS and BMA coefficient estimates and standard errors using data from Easterly-Levine (1997)

	[1]	[2]	[3]	[4]	[5]	[6]
INTERCEPT	-	-	-	-	0.4013	0.1382
	-	-	-	-	0.3985	0.0336
AFRICA	-0.0113	-0.0031	0.9558	0.0761	-	-
	0.0048	0.0053	0.3704	0.0302	-	-
LATINCA	-0.0191	-0.0197	-0.0197	-0.0184	-	-
	0.0036	0.0042	0.0035	0.0037	-	-
DUM60	-0.2657	-0.2200	-0.3643	-0.0028	-	-
	0.0998	0.1765	0.1328	0.0326	-	-
DUM70	-0.2609	-0.2154	-0.3520	0.0009	0.0080	0.0050
	0.0997	0.1745	0.1332	0.0325	0.0134	0.0079
DUM80	-0.2761	-0.2298	-0.3650	-0.0143	-0.0038	-0.0024
	0.0996	0.1751	0.1336	0.0325	0.0132	0.0058
LRGDP	0.0870	0.0756	-0.1090	0.0218	-0.0696	-0.0004
	0.0254	0.0444	0.0986	0.0088	0.1171	0.0027
LRGDPSQ	-0.0063	-0.0056	0.0070	-0.0022	0.0044	-0.0000
	0.0016	0.0029	0.0067	0.0006	0.0088	0.0002
LSCHOOL	0.0117	0.0130	-0.0220	0.0130	-0.0131	-0.0017
	0.0042	0.0056	0.0216	0.0045	0.0194	0.0077
ASSASS	-12.8169	-3.3629	-377.3810	-30.6120	-306.4870	-343.4434
	9.2709	7.8137	165.5661	86.9027	158.4484	181.6948
LLY	0.0162	0.0111	0.1010	0.0129	0.0774	0.0104
	0.0058	0.0083	0.0497	0.0075	0.0483	0.0278
BLCK	-0.0188	-0.0219	-0.0130	-0.0207	-0.0171	-0.0039
	0.0045	0.0053	0.0098	0.0043	0.0107	0.0081
SURP	0.1210	0.1717	0.1200	0.1382	0.1654	0.0948
	0.0314	0.0411	0.0874	0.0357	0.0986	0.1071
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] OLS estimates for model "ALL".
[2] BMA estimates for model "ALL".
[3] OLS 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] BMA 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] OLS on AFRICA subsample.
[6] BMA on AFRICA subsample.

country growth regressions, the performance of sub-Saharan Africa is so much worse than the rest of the world. Rather than remain content with modelling this phenomena as a fixed effect (i.e. a dummy variable) for these countries, Easterly and Levine argue that a major cause of this poor growth performance is the presence of ethnic conflict in these countries. Easterly and Levine construct a measure of ethnic diversity to proxy for this conflict. The magnitude of this variable is substantially higher for sub-Saharan Africa than for the rest of the world.[30] Inclusion of this variable in a cross-country growth regression reduces the size of the African fixed effect and is itself statistically significant. Easterly and Levine conclude

"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." (pg. 1241)

We examine the influence of accounting for model uncertainty on the Easterly and Levine results. Table 1 describes the variables we employ, which are taken from Easterly and Levine (1997). The data are based upon decade-long average observations for the 1960's, 1970's and 1980's, except where indicated in the Table 1 description. We focus on a reexamination of their eq. 3, table IV, which by conventional measures (e.g. the statistical significance of all included variables) is arguably their strongest regression in support of the role of ethnic diversity in growth. This regression is reported in column 1 of our Table 2. The key variable of interest is ELF60, which is a measure of ethnic diversity in each country in 1960.

We explore the role of model uncertainty in two ways. First, we consider the impact of theory uncertainty on inferences concerning the determinants of growth. This is done by constructing a universe of models which consists of all possible combinations of variables that appear in the Easterly and Levine baseline regression. Each of these models is assigned equal probability, which means that each regressor may be thought of as being included in a given regression with probability 1/2; the probability of inclusion of each regressor is, in addition, independent of all others. The subsequent calculation of the posterior mean and standard deviation of each regression

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

coefficient is done using formulas (16) and (17).[31] These results are reported in column 2. 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 -.02 under model averaging, as opposed to the -.017 point estimate reported by Easterly and Levine. Our main conclusion from this exercise is that the Easterly and Levine result is robust to theory uncertainty as we have characterized it.

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. We do this by constructing, for each regressor x_i in the baseline regressors a corresponding variable $x_i\delta_{j,\mathcal{I}_A}$ where $\delta_{j,\mathcal{I}_A} = 1$ if country j is in sub-Saharan Africa, and 0 otherwise. This allows for the possibility that the sub-Saharan African countries possess different parameters than the rest of the world. Column 3 in Table 2 reports 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 regressions 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 the African growth experience as opposed to world growth as a whole. The coefficient estimates for Africa are on the order of 7-10 times greater than the corresponding estimates for the world as a whole.[32] This result is extremely striking and makes clear that the operation of ethnicity on growth is different in Africa, not just the levels of ethnicity. Further, a comparison of the other regressor coefficients for Africa versus the rest of the world makes clear that the African countries should not be treated as partially exchangeable with the rest of the world.

These results in no way diminish the importance of the Easterly-Levine findings. In addition, introduction of potential model uncertainty to check on the

[31] See Table 1 for a list of the countries which are members of sub-Saharan Africa.

[32] The Computational Appendix contains details of the estimation.

robustness of the findings associated with their main equation says nothing about the qualitative evidence they bring to bear. Rather, our results illustrate how additional insights can be achieved through explicitly controlling for model uncertainty.

Finally, we note that this reexamination is quite narrow. A full scale study, in our view, should at a minimum include explicit calculations and presentation of the predictive distribution of the effects of the policy change on growth. Fernandez, Lee, 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.

8. Conclusions

This paper has had two basic aims. First, we attempt to delineate the major criticisms of cross-country growth regressions and to show how to interpret two of these criticisms, theory uncertainty and parameter uncertainty, as violations of a particular assumption – $\mathfrak{F}$ – conditional exchangeability – in the residual components of growth models. Second, we have outlined a framework for conducting and interpreting growth regressions. In terms of conducting regressions, we advocate an explicit modelling of theory and heterogeneity uncertainty and the use of model averaging to condition out strong assumptions. In terms of interpreting regressions, we have argued that the policy objectives associated with a given exercise must be made explicit in the analysis. A decision-theoretic approach to growth regressions has been outlined and its relation to conventional approaches to assessing model robustness has been explored. Finally, in an empirical application, we have shown how attention to model uncertainty can provide new insights into the relationship between ethnicity and growth.

To amplify on 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 which cannot be falsified or confirmed within the statistical procedure that is being employed.[33] Indeed, this is the reason why we have not included a treatment of how to provide more robust arguments in favor of causality in this paper. What we hope is that this paper has provided some initial steps toward the development of a language in which policy relevant empirical growth research may be better expressed.

[33] Similar results are obtained when one compares Africa and the rest of the world. Dropping the African countries from the data set, the OLS estimate for the ELF60 regressor is -.0115 with a standard error of .006; the associated values when averaging is done over different regressor combinations are -.013 and .008. By levels, one would conclude that ethnicity is only marginally statistically significant outside of Africa.

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 which explores only a subset of the model space; the key feature in the design of the algorithm is to ensure that the search proceeds along directions such that it is likely to cover those models which 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, Madigan, Raftery and Volinsky (1999) for additional discussion and implementation. While the reader is asked to see those papers for a full description of the search algorithm, Hoeting, Madigan, Raftery, and Volinsky (1999) 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 $\mathrm{pr}(M_0/D)/\mathrm{pr}(M_1/D)$. 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." (pg. 385).

In implementing the model averaging procedure, this program uses an approximation, due to Raftery (1994), which is based on the idea that, since as the number of observations becomes large, the posterior coefficient distribution will be close to the maximum likelihood estimator, 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 Kadane and Tierney (1986) for technical details. Empirical work by Raftery and coauthors suggests that this approximation works well in practice. Nevertheless, more research is needed on the specification of priors for model averaging; an important recent contribution is Fernandez, Ley and Steel (2000).

References

Aldous, D., (1983), "Exchangeability and Related Topics," in École d'Éte 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," mimeo, Stanford University.

Arnold, S., (1979), "Linear Models with Exchangeably Distributed Errors," Journal of the American Statistical Association, 74, 194-199.

Barro, R., (1991), "Economic Growth in a Cross-Section of Countries," Quarterly Journal of Economics, 106, 2, 407-443.

Barro, R., (1996), "Democracy and Growth," Journal of Economic Growth, 1, 1-27.

Bernard, A. and S. Durlauf, (1996), "Interpreting Tests of the Convergence Hypothesis," Journal of Econometrics, 71, 161-172.

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 in Wiley Classics Library, 2000).

Camerer, C., (1995), "Individual Decision Making," in J. Kagel and A. Roth, eds, The Handbook of Experimental Economics, Princeton: Princeton University Press.

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

Chamberlain, G., (2000), "Econometrics and Decision Theory," Journal of Econometrics, 95, 255-283.

Clark, G., (1987), "Why Isn't the Whole World Developed? Lessons from the Cotton Mills," Journal of Economic History, XLVII, 1, 141-173.

Desdoigts, A., (1999), "Patterns of Economic Development and the Formation of Clubs," Journal of Economic Growth, 4, 3, 305-330.

Doppelhofer, G., R. Miller, and X. Sala-i-Martin, (2000), "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," NBER Working Paper no. 7750.

Draper, D., (1987), "Comment: On the Exchangeability Judgments in Predictive Modeling and the Role of Data in Statistical Research," Statistical Science, 2,4, 454-461.

Draper, D., (1995), "Assessment and Propagation of Model Uncertainty," Journal of the Royal Statistical Society, 57, 1, 45-70.

Draper, D., (1997), "On the Relationship Between Model Uncertainty and Inferential/Predictive Uncertainty," mimeo, School of Mathematical Sciences, University of Bath.

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 Macroeconomics and the Real World, R. Backhouse and A Salanti, eds., Oxford: Oxford University Press.

Durlauf, S. and P. Johnson, (1995), "Multiple Regimes and Cross-Country Growth Behavior," Journal of Applied Econometrics, 10, 365-384.

Durlauf, S., A. Kourtellos, and A. Minkin, (2000), "The Local Solow Growth Model," mimeo, University of Wisconsin at Madison.

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

Easterly, W. and R. Levine, (1997), "Africa's Growth Tragedy: Policies and Ethnic Divisions," Quarterly Journal of Economics, 1203-1250.

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," mimeo, University of Edinburgh.

Fernandez, C., E. Ley, and M. Steel, (2000), "Benchmark Priors for Bayesian Model Averaging," mimeo, University of Edinburgh.

Frankel, J. and D. Romer, "Trade and Growth: An Empirical Investigation," (1996), NBER Working Paper no. 5476.

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

Freedman, D., (1997), "From Association to Causation via Regression," in Causality in Crisis, V. McKim and S. Turner, eds., South Bend: University of Notre Dame Press.

Galor, O., (1996), "Convergence? Inferences from Theoretical Models," Economic

Journal, 106, 1056-1069.

Gilboa, I. and David Schmeidler, (1989), "Maximin Expected Utility with Nonunique Prior," Journal of Mathematical Economics, 18, 141-153.

Goldberger, A., (1991), A Course in Econometrics, Cambridge: Harvard University Press.

Granger, C. and H. Uhlig, (1990), "Reasonable Extreme-Bounds Analysis," Journal of Econometrics, 44, 159-170.

Heckman, J., (2000), "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective," Quarterly Journal of Economics, 45-97.

Hoeting, J., D. Madigan, A. Raftery, and C. Volinsky, (1999), "Bayesian Model Averaging: A Tutorial," Statistical Science, 14, 4, 382-401.

Islam, N., (1995), "Growth Empirics: A Panel Data Approach," Quarterly Journal of Economics, 110, 1127-1170.

Ivanoff, B. and N. Weber, (1996), "Some Characterizations of Partial Exchangeability," Journal of the Australian Mathematical Society, series A, 61, 345-359.

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

Kallenberg, O., (1982), "Characterizations and Embedding Properties in Exchangeability, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 60, 249-281.

Kass, R., and A. Raftery, (1994), "Bayes Factors," Journal of the American Statistical Association, 90, 430, 773-795.

Kourtellos, A., (2000), "A Projection Pursuit Approach to Cross Country Growth Data," mimeo, University of Wisconsin.

Leamer, E., (1983), "Let's Take the Con Out of Econometrics," American Economic Review, 73, 31-43.

Lee, K., M. H. Pesaran, and R. Smith, (1997), "Growth and Convergence in a Multi-Country Stochastic Solow Model," Journal of Applied Econometrics, 12, 4, 357-392.

Levine, R. and D. Renelt, (1992), "A Sensitivity Analysis of Cross-Country Growth Regressions," American Economic Review, 82, 942-963.

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, 1535-1546.

Mallows, C., (1998), "The Zeroth Problem," The 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, CVII, 407-437.

McAleer, M., A. Pagan, and P. Volker, (1985), "What Will Take the Con Out of Econometrics?," American Economic Review, 75, 3, 293-307.

Pack, H., (1994), "Endogenous Growth Theory: Intellectual Appeal and Empirical Shortcomings," Journal of Economic Perspectives, 8, 1, 55-72.

Prescott, E., (1998), "Needed: A Theory of Total Factor Productivity," International Economic Review, 39, 525-552.

Quah, D., (1996a), "Convergence Empirics Across Economies with (Some) Capital Mobility," Journal of Economic Growth, 1, 95-124.

Quah, D., (1996b), "Empirics for Growth and Economic Convergence," European Economic Review, 40, 1353-1375.

Quah, D., (1997), "Empirics for Growth and Distribution: Polarization, Stratification, and Convergence Clubs," Journal of Economic Growth, 2, 27-59.

Pritchett, L., (2000), "Patterns of Economic Growth: Hills, Plateaus, Mountains, and Plains," mimeo, World Bank.

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

Raftery, A., D. Madigan, and J. Hoeting, (1997), "Bayesian Model Averaging for Linear Regression Models," Journal of the American Statistical Association, 92, 437, 179-191.

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

Romer, P., (1986), "Increasing Returns and Long Run Growth," Journal of Political Economy, 94, 1002-1037.

Romer, P., (1990), "Endogenous Technical Change," Journal of Political Economy, 98, (5 Part 2), 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, 307-344.

Schultz, T. P., (1999), "Health and Schooling Investments in Africa," Journal of Economic Growth, 13, 3, 67-88.

Sims, C., (1980), "Macroeconomics and Reality," Econometrica, 48, 1, 1-48.

Stewart, M., (1984), "Significance Tests in the Presence of Model Uncertainty and Specification Search," Economics Letters, 16, 309-313.

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

Wasserman, L., (1997), "Bayesian Model Selection and Model Averaging," Technical Report 666, Department of Statistics, Carnegie Mellon University.

Wolcott, S. and G. Clark, (1999), "Why Nations Fail: Managerial Decisions and Performance in Indian Cotton Textiles, 1890-1938," Journal of Economic History, 59, 2, 397-423.