Growth Regressions and Policy Evaluation

Philippe Aghion and Steven Durlauf

March 31, 2007

Preliminary and Incomplete
Please do not Circulate
Comments Welcome

Prepared for the World Bank Commission on Growth. Financial support from the National Science Foundation is gratefully acknowledged.

1. Introduction

In this paper, we discuss the relationship between growth regressions and growth policy evaluation. Our objectives are twofold. First, we describe some general difficulties that exist in translating growth regressions into policy. Second, we propose some ways of rendering the use of growth regressions more policy relevant. We fully recognize that actual policy recommendations depend on much more than the output of regression exercises. Nevertheless, such exercises are often a source of support for policy claims.

Our argument is that while much current empirical practice is inadequate to the task of policy analysis, this need not be the case. From the perspective of policy evaluation, current econometric practice in growth analysis is often frustrating as one typically sees a focus on the question of the identification of a best fitting model for a particular statistical model for a given data set, as opposed to the construction of a description of the uncertainty associated with the choice of a given policy. We therefore describe some ways in which growth regressions can be made more "policy-relevant." Our discussion in essence places relatively standard ideas from statistical decision theory into the growth regression context.

Our focus in this paper is on conceptual issues involved in the interpretation of growth regressions for policy comparison. In our companion paper (Aghion and Durlauf (2007)), we discuss questions of causality versus correlation, the role of fixed effects and the like. These considerations relate to the question of interpretation of regressions with respect to their ability to provide the sorts of counterfactual analyses needed for growth regression. Here we assume that a correctly specified growth model can be interpreted causally and address the question of how it can be brought to data in a way that makes its use most policy relevant.

The relationship between decisions and data analysis is of course of longstanding importance in the statistics literature. In many respects, our arguments are no more than simple versions of the ideas pioneered by Abraham Wald and Leonard J. Savage, only applied to the growth context. We mention this distinguished pedigree not to puff up our arguments, but rather to note that there is a deep set of ideas on linking data to decisions.

2. Background: the canonical regression

In the specific context of growth regressions, policy evaluation is commonly done on the basis of assessing the statistical significance of a regression parameter corresponding to the policy of interest (in the case below, δ). A typical growth regression has the structure

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

where g_i is real per capita growth across some fixed time interval, X_i is a set of regressors suggested by the Solow growth model (population growth, technological change, physical and human capital savings rates transformed in ways implied by the model), Z_i is a set of additional control variables suggested by new growth theories, p_i is the policy variable of interest, and ε_i is an error. The distinction between X_i and Z_i is important in econometric practice because while X_i variables are generally constant across empirical studies, there is relatively little consensus on which Z_i variables should be included. We focus on cross-section regressions rather than panel regressions for ease of exposition. None of our main messages change qualitatively when one uses panel data.

In our view, the way in which analyses focus on δ in order to evaluate growth policies is flawed in two respects. First, much empirical growth analysis fails to capture the appropriate level of uncertainty associated with estimates of δ. Second, empirical analyses typically do not formulate their analyses in decision-theoretic terms, so that the analyses fail to appropriately describe the consequences of policies.

2. model uncertainty

The researcher's uncertainty about the value uncertainty of δ exists at distinct two levels. One level is the uncertainty associated with the parameter, conditional on a choice of (1). This level of uncertainty is of course assessed in virtually every empirical study. What is not fully assessed is the uncertainty associated with the specification (1). It is typical for a given paper that the specification of (1) is taken as essentially known; while some variations of a baseline model are often reported, via different choices of elements of Z, standard empirical practice does not systematically account for the sensitivity of claims about δ to model choice. This model uncertainty is reflected in the fact that, as documented in Durlauf, Johnson, and Temple (2005), there are over 40 general growth theories and 130 specific growth determinants which have appeared in various cross country regressions, as well as numerous formulations of alternatives to (1) which allow for alternative functional specifications.

The existing body of empirical growth regressions, in our view, suggests 4 general levels of model uncertainty that need to be systematically addressed in each exercise. To be clear, we do not claim that this decomposition of model uncertainty is unique; our divisions are not natural kinds. Rather, we find these levels useful in thinking about where empirical claims about a given regression may go awry.

First, there is theory uncertainty. Theories about the roles of institutions, geography, ethnic heterogeneity, etc. in growth find statistical analogs in particular constructions of growth regressors. As a result, one finds different estimates of the effects of a given policy because of different stances on which growth theories have empirical salience and so must be controlled for in a regression.

Second, there is measurement uncertainty. Growth theories often do not translate into natural empirical analogs. If one makes an argument that property rights affect growth, this does not lead to any particular implications about which measures of property rights ought to be included. One thus finds a range of different choices of control variables that amount to differences in ways to empirically proxy for a given theory.

Third, there is specification uncertainty. The linear specification (1) is not a natural one for much of growth analysis. The standard justifications of the specification are that it either represents a linear approximation to the growth process associated with the Solow model or is the law of motion under a Cobb-Douglas aggregate production function (cf. Barro and Sala-i-Martin (1995) and Mankiw, Romer and Weil (1992)); but this begs the question of how to understand the results of growth regressions when these justifications are false.[1] Perhaps unsurprisingly, the application of a linear growth model to a nonlinear environment can produce very misleading conclusions. This is easily seen for threshold externality models of the type described by Azariadis and Drazen (1990). In these models, one finds multiple steady states, each of which has the property that local to it, the Solow approximation is appropriate. As shown by Bernard and Durlauf (1996), when one applies a linear regression to data generating from this environment, one can find statistical evidence of convergence even though countries are associated with different steady states. Another type of nonlinearity concerns interactions between variables. An example of where this matters is the study of the relationship between foreign aid and growth, cf. Burnside and Dollar (2000). Here, the policy question concerns whether the effects of aid on growth depend on the quality of a country's government. While their specific findings have been challenged, the methodology reflects how nonlinearity naturally arises in growth/policy contexts.

Specification uncertainty also exists with respect to transition versus steady state dynamics. An example of this is the dependence of parameter values on current state variables. Lucas (2000) argues that claims of substantial international income inequality over the long run, based on the iteration of time invariant Markov processes for the contemporaneous income distribution are flawed as they do not account for the possibility that these parameters are evolving. Put differently, he argues in favor of a model where the current income distribution reflects transition dynamics whereas studies such as Jones (1997) assume that data under study are draws from an invariant Markov process. Substantively, Lucas argues that there will be substantially greater convergence than predicted by the time invariant process in the world income distribution because of transition of poorer economies. This distinction between transition and steady state behavior can have the opposite implications in other contexts. Bernard and Durlauf (1996) argue that convergence tests for developed economies need to account for the

[1]See Duffy and Papageorgiou (2000) for evidence against the Cobb-Douglas assumption.

distinction between transitory and steady state behavior, and so argue that some evidence of convergence is flawed in that it is consistent with the presence of multiples steady states, including poverty traps. For our purposes what matters is that this is another form of specification uncertainty, in this case with respect to the assumptions made about the stationarity of the data.

These different types of model uncertainty matter because, while there are differing degrees of evidentiary support for the various models, economic theory does not provide much guidance in terms of the choice of a particular model. In other words, relative to a space of possible growth models suggested by modern growth theory, the focus on a single candidate, or small set of candidates similar to a baseline model represents the imposition of very strong prior information on the part of the researcher. As argued by Brock and Durlauf (2001), such strong prior information seems hard to defend given the openended nature of growth theories, i.e. their mutual logical compatibility.

2. statistical decision theory and empirical practice

The second limitation of current empirical practice for policy evaluation concerns the lack of explicit attention to policy evaluation as a decision problem. Specifically, it is relatively uncommon to use growth regressions to explicitly inform policy analysis, as opposed to using statistical significance of a policy variable as evidence in favor of a policy option. To see the difference, suppose that a policymaker interestiug in affecting growth rate in country i is choosing between policy variable at some fixed level $\overline{p}$ and an alternative setting at $\overline{\overline{p}}$. Suppose that the policymaker has a payoff function

$$V\left(g_i,\middle|D_i\,,p_i\right). \tag{2}$$

The conditioning on D_i allows for country-specific heterogeneity. The policy evaluation under an expected payoff calculation will amount to computing

$$EV\left(g_i,\middle|D_i\,,\overline{\overline{p}}\right)-EV\left(g_i,\middle|D_i\,,\overline{p}\right) \tag{3}$$

where $E\left(\cdot\right)$ is an expected value operator. What matters for our purposes is that (3) does not correspond to a test of statistical significance of the policy parameter. For it to do so, it would be necessary (assuming that statistical significance is based on a $t-$statistic) that

$$\begin{gathered} EV\left(g_i,\middle|D_i\,,\overline{\overline{p}}\right)-EV\left(g_i,\middle|D_i\,,\overline{p}\right)= \\ \hat{\delta}\left(\overline{\overline{p}}-\overline{p}\right)-\kappa var\left(\hat{\delta}\left(\overline{\overline{p}}-\overline{p}\right)\middle|\delta=0\right)^{1/2}>0, \end{gathered} \tag{4}$$

which of course reduces to

$$\hat{\delta}-\kappa var\left(\hat{\delta}\middle|\delta=0\right)^{1/2}>0, \tag{5}$$

where κ is the significance level. This is a very special case and means that the policymaker evaluates policies by their effect on the component of growth associated with the policy, i.e. δp_i, rather than the effect of the control variable on growth the growth process per se. We are unaware of any argument that such preferences are sensible ones.[2]

Does the fact that tests of statistical significance do not correspond to policy comparisons matter in practice? It is easy to see how explicit payoff comparisons can lead to very different assessments than those implied by statistical significance. Suppose that the question of interest whether to choose $\overline{p}$ or $\overline{\overline{p}}$, assume $\overline{p}<\overline{\overline{p}}$.. Abstract away

[2]Note that the absence of sensible preferences to produce this representation is *not* a function of the fact that t-statistics, as frequentist objects, focus on the probability description of an observable (the normalized parameter estimate) conditional on an unknown value (i.e. 0) whereas Bayesian approaches focus on the probability description of an unobservable, the policy parameter, given the observed data.

all issues of model uncertainty and the Lucas critique[3], i.e. assume that (1) is a structural model so that parameters do not depend on the policy levels. Further assume that the various growth determinants in (1) are nonstochastic and uncorrelated within the sample with each other.[4] The loss function associated with the $t-$statistic rule will compute (5), which under our orthogonality assumptions, reduces to

$$\hat{\delta} - \frac{\kappa}{N^{1/2}} \left(\frac{\widehat{\text{var}}\, \varepsilon}{\widehat{\text{var}}\, p} \right)^{1/2} > 0 . \tag{6}$$

Here $\widehat{\text{var}}\, p$ and $\widehat{\text{var}}\, \varepsilon$ denote the sample variances of the respective variables. The policy recommendation will not depend on the country considered since it depends on inferences about a parameter that is constant across countries.

In contrast, suppose that the expected payoff is determined by the mean and variance of the growth level. Specifically, suppose that the expect payoff has the form

$$EV\left(g_i, \middle| D_i, p_i\right) = E\left(g_i \middle| D_i, p_i\right) - \kappa\, \text{var}\left(g_i \middle| D_i, p_i\right)^{1/2} . \tag{7}$$

(We choose this representation as it best mimics the structure of the $t-$statistic rule.) In this case, the payoff as a function of policy will equal

$$\begin{gathered} V\left(g_i, \middle| D_i, p_i\right) \approx \\ \hat{\alpha} + X_i \hat{\beta} + Z_i \hat{\gamma} + p_i \hat{\delta} - \\ \kappa \left(\sum x_{i,k}^2 \left(\frac{\widehat{\text{var}}\, \varepsilon}{N \widehat{\text{var}}\, x_k} \right) + \sum z_{i,l}^2 \left(\frac{\widehat{\text{var}}\, \varepsilon}{N \widehat{\text{var}}\, z_l} \right) + p_i^2 \left(\frac{\widehat{\text{var}}\, \varepsilon}{N \widehat{\text{var}}\, p} \right) + \widehat{\text{var}}\, \varepsilon \right)^{1/2} . \end{gathered} \tag{8}$$

[3]We address model uncertainty below. The Lucas Critique is typically (and unfortunately) ignored in the empirical growth literature.

[4]We make these assumptions to simplify calculations.

(The approximation in the formula follows from the use of sample moments in place of posterior density moments.) For our special example of orthogonal regressors, the policy change is justified if

$$
\begin{aligned}
& EV\left(g_i,\middle|D_i, \bar{\bar{p}}\right) - EV\left(g_i,\middle|D_i, \bar{p}\right) \approx \\
& \hat{\delta} - \\
& \kappa\left(\sum x_{i,k}^2\left(\frac{\widehat{\mathrm{var}}\,\varepsilon}{N\,\widehat{\mathrm{var}}\,x_k}\right) + \sum z_{i,l}^2\left(\frac{\widehat{\mathrm{var}}\,\varepsilon}{N\,\widehat{\mathrm{var}}\,z_l}\right) + \bar{\bar{p}}^2\left(\frac{\widehat{\mathrm{var}}\,\varepsilon}{N\,\widehat{\mathrm{var}}\,p}\right) + \widehat{\mathrm{var}}\,\varepsilon\right)^{1/2} + \\
& \kappa\left(\sum x_{i,k}^2\left(\frac{\widehat{\mathrm{var}}\,\varepsilon}{N\,\widehat{\mathrm{var}}\,x_k}\right) + \sum z_{i,l}^2\left(\frac{\widehat{\mathrm{var}}\,\varepsilon}{N\,\widehat{\mathrm{var}}\,z_l}\right) + \bar{p}^2\left(\frac{\widehat{\mathrm{var}}\,\varepsilon}{N\,\widehat{\mathrm{var}}\,p}\right) + \widehat{\mathrm{var}}\,\varepsilon\right)^{1/2} \\
& > 0.
\end{aligned}
\tag{9}
$$

For our purposes, there are two main points in contrasting policy evaluation based on statistical significance, as done in (6) and policy evaluation based on (9). First, the choice of p_i will depend on the uncertainty associated with each of the parameters associated with the growth determinants, not just the uncertainty associated with δ. Second, the choice will depend on the individual (i.e. country-specific) values of the various growth determinants. Together, these factors mean that, since different countries will have different levels of uncertainty associated with the growth process, there is no reason why growth recommendations need to be common across countries. Notice that for the orthogonal determinants example, it is easy to see how greater uncertainty associated with the various growth components of a country directly affect the calculation of the marginal “cost” with respect to the variance of the growth process of an increase in p_t.

Brock, Durlauf, and West (2003) apply this type of calculation to the question of lowering tariffs in sub-Saharan African countries, concluding that while the policy is generally to be recommended, it is not uniformly so. Brock, Durlauf, and West (2003) further found that the changes in the variance of the growth rate associated with a one unit change in tariffs was of an order of magnitude smaller than the uncertainty associated with the parameter estimate itself, so that the $t-$statistic rule gave very misleading answers as to the payoff-relevant uncertainty induced by the proposed policy change.

This comparison between the effects of a policy change on the policy component to growth, versus the level of growth, is itself problematic as the latter formulation begs the question of the appropriate payoff function. Presumably, a policymaker is interested in the sequence of output levels in a country rather than one period growth per se. This is where the focus on cross-sectional analysis ignores fundamental questions of policy evaluation. And of course, deeper issues of multiple objectives also exist.

3. Suggestions for empirical practice

i. model uncertainty

How can one constructively address model uncertainty? One approach has been developed in work by Brock and Durlauf (2001), Brock, Durlauf and West (2003), Fernandez, Ley and Steel (2001) and Doppelhofer, Miller Sala-i-Martin (2004) and uses model averaging techniques. This research develops inferences about growth determinants based on a model space M rather than a single model m. Other applications in growth contexts include Cohen-Cole, Durlauf and Rondina (2006), Jones and Schneider (2006) and Masanjala and Papageorgiou (2005).

In implementing model averaging, the construction of the model space is arguably the most difficult step. M embodies the researcher's judgments as to what models are appropriate to consider and so the construction of M cannot be reduced to an algorithm. That said, one rule we do believe is appropriate is that the model space should span all coherent models that may be produced from those elements it contains. For example, if one growth regression includes geographical variables as growth determinants, and a second contains institutional variables, then there should be a model in M that contains the union of these variables. From the perspective of variable selection, this notion of spanning the space is thus straightforward. For this reason, it is convenient to interpret differences between models as differences in choices of control variables. To do this, nonlinearities are treated as products of variables, e.g. an indicator function with respect to some level of initial income multiplied by the level of human capital savings can capture an Azariadis-Drazen type of threshold externality.

For each model m in the space, one assigns a prior probability $\mu(m)$ which allows computation of a posterior model probability $\mu(m|D)$, where D constitutes the data available to the analyst. These posterior probabilities in essence allow the relative goodness of fit of each model to affect the relative model weights. For a Bayesian, each model produces a posterior probability density $\mu(\delta|D,m)$. The model averaged posterior probability is thus

$$\mu(\delta|D) = \sum_{m \in M} \mu(\delta|D,m)\,\mu(m|D). \tag{10}$$

As shown by Leamer (1978), the first two moments of this posterior density are

$$E(\delta|D) = \sum_{m \in M} E(\delta|D,m)\,\mu(m|D), \tag{11}$$

and

$$var(\delta|D) = \sum_{m} var(\delta|D,m)\mu(m|D) + \sum_{m}\left(E(\delta|D) - E(\delta|D,m)\right)^2 \mu(m|D). \tag{12}$$

One can also implement model averaging in frequentist contexts, so long as one is willing to take a stand on the calculation of the posterior model probabilities. Using frequentist estimators, one computes the model-specific estimates $\hat{\delta}_m$ and $\text{var}(\hat{\delta}_m|\delta)$. One then combines the information about the policy parameter across the model space to produce a model averaged estimate of the policy parameter

$$\hat{\delta}_M = \sum_{m \in M} \hat{\delta}_m \mu\left(m|D\right), \tag{13}$$

and associated model averaged estimate of its variance

$$var\left(\hat{\delta}_M \,|\delta\right) = \sum_m var\left(\hat{\delta}_m \,|\delta\right)\mu\left(m|D\right) + \sum_m \left(\hat{\delta}_M - \hat{\delta}_m\right)^2 \mu\left(m|D\right). \tag{14}$$

The term $\sum_m \left(E\left(\delta|D\right) - E\left(\delta|D,m\right)\right)^2 \mu\left(m|D\right)$ in the Bayesian case and the term $\sum_m \left(\hat{\delta}_M - \hat{\delta}_m\right)^2 \mu\left(m|D\right)$ in the frequentist case capture how the uncertainty associated with an estimate of a policy effect changes when accounts for model uncertainty. These terms reflect the extra variance associated with the parameter that is attributable to its fluctuation across models. Draper (1995) is a valuable discussion on how the failure to account for this term can lead to spurious claims about forecast accuracy. His arguments apply equally well to evaluating policy counterfactuals.

What implications does a systematic evaluation of model uncertainty have for empirical claims that have been made on the basis of growth regressions? The two broad ranging efforts to account for model uncertainty in identifying salient growth determinants are Fernandez, Ley and Steel (2001) and Doppelhofer, Miller Sala-i-Martin (2004). In these papers, a large number of growth regressors are gathered, various combinations are constructed to define a model space, and analyses are conducted to identify robust growth determinants. Using the standard that a variable is included in the "true specification" with posterior probability .9 or greater, Doppelhofer, Miller and Sala-i-Martin (2004) working with 31 potential growth determinants find four such variables: initial income, fraction of GDP in mining, number of years the economy has been open, and fraction of the population following Confucianism. Fernandez et al. also find four robust determinants: initial income, fraction of the population following Confucianism, life expectancy, and rate of equipment investment. From the policy perspective, this leads to the unsurprising conclusions that capital accumulation and international trade are useful to growth. Perhaps more interesting, Doppelhofer, Miller, and Sala-i-Martin

(2004) failed to find any evidence that policy variables such as government consumption, civil liberties, public education spending, or war participation affect growth; Fernandez et al find no evidence for these variables either. From the perspective of these exercises, cross country growth regressions have little to say on policy.

One reason why the evidence of robust growth determinants is so scanty is the following. Recall that posterior model probabilities have the structure

$$\mu(m|D) \propto \mu(m)\mu(D|m), \tag{15}$$

where $\mu(D|m)$ is the likelihood of the data given a model and $\mu(m)$ is the prior. While the formulation of the likelihood is technically straightforward, the specification of the prior is problematic. Both Fernandez, Ley and Steel (2001) and Doppelhofer, Miller Sala-i-Martin (2004) assume that each growth determinant enters into the correct growth model independently from the others. This assumption is not tenable. Since many of the candidate regressors are simply different proxies for the same underlying growth explanation. As these regressors are highly collinear, it is easy to see why it would be difficult to establish evidence in favor of one in the presence of the others.

There is no accepted solution to the problem of assigning priors to model spaces. Brock, Durlauf, and West (2003) propose a hierarchical prior structure that assigns probabilities to individual regressors based on the levels of model uncertainty we have described. This is illustrated in Figure 1, which is a slight modification of a figure in Cohen-Cole, Durlauf and Rondina (2006). There are connections between the hierarchical prior structure and the red bus/blue bus problem in discrete choice theory, see the discussion in Brock and Durlauf (2001); Brock, Durlauf and West use these to create nested model probabilities so that, for example, the weight given to a theory is not artificially increased by an abundance of empirical proxies for it. Another approach for assigning priors in model spaces is due to Edward George and employed in a growth context by Durlauf, Kourtellos, and Tan (2006). This approach is based on the construction of a dilution prior, in which probability mass in the model space accounts for model similarity. This may be understood as assigning probabilities in a way that the addition of a model that to the space that is identical to a previous one would simply mean that the two equal models divide the probability assigned to the first. Yet another approach has been proposed by Doppelhofer and Weeks (2007) who consider ways to identify whether variables tend to appear together or separately in growth regressions. Here the emphasis is less on prior construction than enriching the set of questions one asks.

As there does not exist a clear best practice for prior construction, we think that good empirical work should at a minimum verify that claims are robust to different choices of priors among those that have functioned as standards within the literature. Further, we believe it is important to report results for some special cases where priors are degenerate, specifically, when one assigns probability 1 to theories of particular interest. For example, reporting analyses when a prior of 1 is assigned to the neoclassical model and a prior of 1 is assigned to an endogenous innovations model is useful in communicating to a policymaker the consequences of taking a theoretical stance. While a policymaker presumably has little attachment to particular choices of variable measurement and model specification (as we have defined it), it is natural that policymakers have strong priors with respect to growth theories.

We also envision a role for evaluating policies when one does not want to deal with priors. This leads to the consideration of minimax and minimax regret approaches to policy evaluation. Hansen and Sargent (see (2007) for a brilliant and exhaustive treatment) have pioneered the application of minimax analysis in macroeconomic analysis. In our context, minimax would amount to evaluating policies on the basis of of its model-specific expected payoff for the least favorable model in the model space. Brock, Durlauf, West and Rondina (2007) argue in favor of minimax regret for policy comparisons when a researcher chooses not to assign priors.[5] [6] What matters for our purposes is that both minimax and minimax regret allow for policy evaluation when a

[5]See Manski (2005,2006) for discussion of minimax regret in microeconomic policy contexts.

[6]One issue in the use of minimax regret in policy analysis is that the criterion does not obey independence of irrelevant alternatives, something noted by Chernoff (1954). Brock, Durlauf, Nason and Rondina (2007) argue this failure should not disqualify the criterion from use.

researcher does not want to take a stance on model priors. Both of these approaches are explicitly decision-theoretic, which is the area we turn to next.

ii. decision-relevant empirical analysis

If the goal of an empirical exercise is to evaluate expected payoffs under policies and the payoff function is known, then it is conceptually simple to account for model uncertainty. One engages in model averaging to compute posterior densities for growth under alternative policies and computes the associated expected values. In fact, from the perspective of policy evaluation, there is an important sense in which the model is "uninteresting" as it is simply an unobservable that needs to be integrated out of the expected payoff calculation. What we mean by this is the following. For a given model m, and space of possible policies P, the optimal policy problem is

$$\max_{p \in P} \int V\left(g_i, D_i \middle| D, \bar{\bar{p}}\right) \mu\left(g_i \middle| D, m\right) \tag{16}$$

whereas if model uncertainty is present, the associated policy problem is

$$\max_{p \in P} \int V\left(g_i, D_i \middle| D, \bar{\bar{p}}\right) \sum_{m \in M} \mu\left(g_i \middle| D, m\right) \mu\left(m \middle| D\right) \tag{17}$$

where D denotes the data available across all data periods. This approach is taken, for example, in Brock, Durlauf, and West (2003), who report country-specific estimates of the effects on the growth rates of a tariff reduction for sub-Saharan African countries.

A full consideration of model uncertainty should, in our view, also account for uncertainty about the payoff function as well as uncertainty about the data generating function, which is the sort of model uncertainty we have so far discussed. David Cox (2006, p. 163) in fact critiques statistical decision theory for this reason:

Wald's treatment of decision theory supposed that a utility function is available but not a prior is available; in some respects this is often an odd specification in that the choice of the utility may be at least as contentious as that of the prior.

Unlike uncertainty about the data generating process, much less is known about the ways to proceed when one does not know the correct payoff function.

In the growth context, we see obvious places where uncertainty about the payoff function matters. Most obvious, inferences about payoffs will be sensitive to the way in which risk aversion is embodied in the loss function. This is evident when one reexamines the loss function (7) and considers the question of the interaction of the effects of a policy change on the variance term when the coefficient on the term is changed from ½ to 1. This problem is especially important since growth models generally do not have a sufficiently rich structure to allow one to calculate aggregate agent utility, as may be done in dynamic stochastic general equilibrium models (DSGE) of the type now used in business cycle analysis. In other words, we do not believe that one can appeal to the typical current growth model to provide welfare calculations in the way that a DSGE model allows.

Further, there is scope to consider a larger set of arguments to the payoff function than simply growth; an obvious alternative candidate is some notion of equality.[7] While we do not take a stance on the appropriate measure of equality, we do note that in order to assess growth/equality tradeoffs (should they exist) using standard decision-theoretic methods it is necessary to take a stance on the loss function. Since the choice of such a function strikes us as problematic, we think it is worthwhile to develop a notion of robustness for the case where the payoff function is not known. This view of robustness will often lead to an interest in policies that tend to avoid large tradeoffs. Work of this type is pursued in the very different context of monetary policy evaluation (where the

[7]Here we are following Amartya Sen (e.g. Sen (1985)) in allowing the evaluation of the consequences of policies by more than their utility implications. For growth contexts, equality measures (primarily associated with mobility) seem to be natural candidates. Hence tradeoffs are possible between policymaker objectives, something that is not naturally allowed in nonconsequentialist, specifically deontological, considerations. We are thus allowing for a policymaker to value equality outside of its role, via the concavity of utility functions, in maximizing aggregate utility.

tradeoffs involve aggregate volatility at different business cycle frequencies) in Brock, Durlauf, and Rondina (2007).

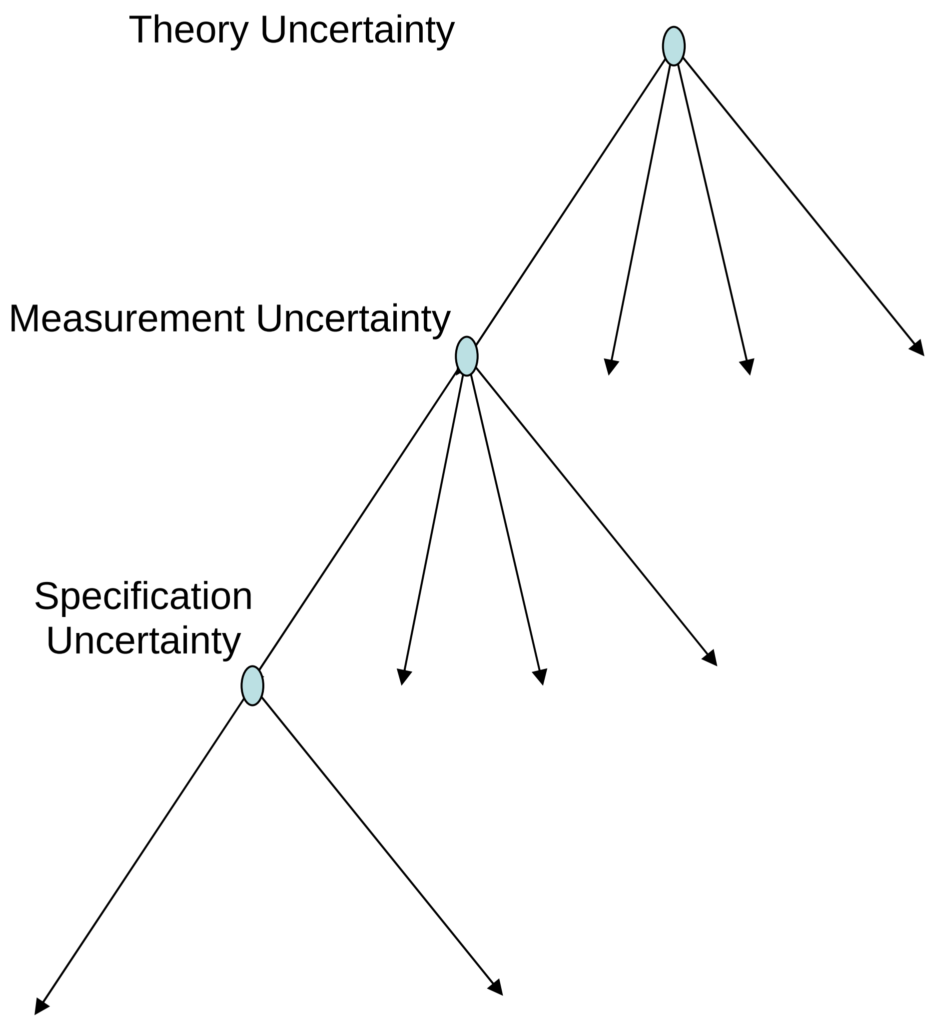
The development of appropriate ways to think about policy robustness strikes us an important area for growth research. The Hansen/Sargent program has been immensely successful in macroeconomics and suggests that the development of an analog in growth is feasible. Growth analyses, because they deal model uncertainty of a somewhat different type with respect to the data generating process and because of the addition of payoff function uncertainty, will require the development of somewhat distinct, albeit complementary, tools from those developed by Hansen and Sargent.

We would also note that robustness issues seem salient for microeconomic issues such as development assistance. Many of Easterly's (2001,2006) critiques of international aid and the third world can be cast in terms of the failure to appropriately account for model uncertainty and to use an appropriate decision theoretic language for evaluating programs.

4. Conclusions

Growth regressions can play a useful role in policy assessment, so long as this role is appropriately tailored to the question of a decision problem. For policy assessment, the determination of the correct growth model has only instrumental value. The appropriate object of interest appears to be posterior densities of the variables of interest under alternative policy choices. These densities should be constructed in a way that fully reflects how little consensus there is among growth theorists as to the relative salience of various theories as well as the lack of strong guidance from theory towards the specification of a particular statistical model. Evaluation of these posterior densities should further reflect explicit consideration of a policymaker's objective function and should allow for uncertainty as to the appropriate choice of this function. While there are no panaceas, we think the common-sense suggestions we have outlined can help to increase the credibility of growth regressions as a source of evidence in favor of one policy or another.

Figure 1
Model Uncertainty in Growth Models



Bibliography

Azariadis, C. and A. Drazen (1990), “Threshold Externalities in Economic Development”, Quarterly Journal of Economics, 105, 2, 501-526.

Barro, R. and X. Sala-i-Martin (1995), Economic Growth (McGraw-Hill, New York).

Bernard, A. and S. Durlauf (1996), “Interpreting Tests of the Convergence Hypothesis”, Journal of Econometrics 71, 1-2, 161-73.

Brock, W. and S. Durlauf (2001), “Growth Empirics and Reality”, World Bank Economic Review 15, 2, 229-272.

Brock, W., S. Durlauf and G. Rondina, (2007), “Design Limits and Dynamic Policy Analysis,” mimeo.

Brock, W., S. Durlauf and K. West (2003), “Policy Evaluation in Uncertain Economic Environments (with discussion)”, Brookings Papers on Economic Activity 1:235-322.

Burnside, C. and D. Dollar, (2000), “Aid, Policies, and Growth, *American Economic Review*, 90, 4, 847-868.

Cohen-Cole, E., S. Durlauf and G. Rondina, (2005), “Nonlinearities in Growth: From Evidence to Policy,” mimeo.

Cox, D., (2006), *Principles of Statistical Inference*, Cambridge: Cambridge University Press.

Doppelhofer, G., R. Miller and X. Sala-i-Martin (2004), “Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach”, American Economic Review 94(4):813-835.

Doppelhofer, G. and M. Weeks, (2007), “Jointness of Growth Determinants, mimeo.

Draper, D. (1995), “Assessment and Propagation of Model Uncertainty”, Journal of the Royal Statistical Society, series B, 57:45-70.

Duffy, J. and C. Papageorgiou (2000), “A Cross-Country Empirical Investigation of the Aggregate Production Function Specification,” *Journal of Economic Growth*, 5,87-120.

Durlauf, S., P. Johnson, and J. Temple, (2005), “Growth Econometrics,” in *Handbook of Economic Growth*, P. Aghion and S. Durlauf eds., Amsterdam: North-Holland.

Easterly, W., (2001), *The Elusive Quest for Economic Growth*, Cambridge: MIT Press.

Easterly W., (2006), *The White Man's Burden,* New York: Penguin.

Fernandez, C., E. Ley and M. Steel (2001), "Model Uncertainty in Cross-Country Growth Regressions", Journal of Applied Econometrics 16(5):563-76.

Hansen, L. and T. Sargent, (2007), Robsutness, manuscript in progress, Princeton University Press.

Jones, C., (1997), "On the Evolution of the World Income Distribution," *Journal of Economic Perspectives*, 11, 19-36.

Jones, G. and W. Schneider, (2006) "Intelligence, Human Capital, and Economic Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *Journal of Economic Growth*, 11, 1, 71-93.

Lucas, R., (2000), "Some Macroeconomics for the 21st Century," *Journal of Economic Perspectives*, 1, 159-168.

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

Manski, C., (2005), "Statistical Treatment Rules for Heterogeneous Populations," *Econometrica*, 72, 1221-1246.

Manski, C., (2006), "Minimax-Regret Treatment Choice with Missing Outcome Data," *Journal of Econometrics*, forthcoming.

Masanjala, W. and C. Papageorgiou, (2005), "Rough and Lonely Road to Prosperity: A Reexamination of the Sources of Growth in Africa using Bayesian Model Averaging," mimeo, IMF.

Sen, A., (1985), "Well Being, Agency and Freedom," *Journal of Philosophy*, 84, 2, 169-221.