NBER WORKING PAPER SERIES

POLICY EVALUATION IN UNCERTAIN ECONOMIC ENVIRONMENTS

William A. Brock
Steven N. Durlauf
Kenneth D. West

Policy Evaluation in Uncertain Economic Environments
William A. Brock, Steven N. Durlauf, and Kenneth D. West
NBER Working Paper No. 10025
October 2003
JEL No. C10, C50, E52, O40

ABSTRACT

This paper develops a decision-theoretic approach to policy analysis. We argue that policy evaluation should be conducted on the basis of two factors: the policymaker's preferences, and the conditional distribution of the outcomes of interest given a policy and available information. From this perspective, the common practice of conditioning on a particular model is often inappropriate, since model uncertainty is an important element of policy evaluation. We advocate the use of model averaging to account for model uncertainty and show how it may be applied to policy evaluation exercises. We illustrate our approach with applications to monetary policy and to growth policy.

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

Kenneth D. West
Department of Economics
University of Wisconsin
1180 Observatory Drive
Madison, WI 53706-1393
and NBER
kdwest@wisc.edu

The number of separate variables which in any particular social phenomenon will determine the result of a given change will as a rule be far too large for any human mind to master and manipulate them effectively. In consequence, our knowledge of the principle by which these phenomena are produced will rarely if ever enable us to predict the precise result of any concrete situation. While we can explain the principle on which certain phenomena are produced and can from this knowledge exclude the possibility of certain results…our knowledge will in a sense only be negative, i.e. it… will not enable us to narrow the range of possibilities sufficiently so that only one remains.

Friedrich von Hayek[2]

It will be remembered that the seventy translators of the Septuagint were shut up in seventy separate rooms with the Hebrew text and brought with them, when they emerged, seventy identical translations. Would the same miracle be vouchsafed if seventy multiple correlators were shut up with the same statistical material? And anyhow, I suppose, if each had a different economist perched on his *a priori*, that would make a difference to the outcome.

John Maynard Keynes[3]

1. Introduction

This paper describes some approaches to macroeconomic policy evaluation in the presence of uncertainty about the structure of the environment under study. The perspective we discuss is designed to facilitate policy evaluation for several forms of uncertainty. For example, our approach may be used when an analyst is unsure about the appropriate economic theory that should be assumed; it may also be employed when an analyst is unsure about the particular functional forms that translate a general theory into a form amenable to statistical analysis. As such, these methods are, we believe, particularly useful in a range of macroeconomic contexts where there are fundamental disagreements as to the determinants of the problem under study. In addition, this approach recognizes that even if one agrees on the underlying economic theory that describes a phenomenon, policy evaluation often requires taking a stance on details of the economic environment such as lag lengths and functional form that are not specified by

the theory. As such, our analysis is motivated by similar concerns as led to the development of model calibration methods. Unlike the usual calibration approach, we do not reject formal statistical inference methods but rather incorporate model uncertainty into them.

The key intuition underlying our analysis is that for a broad range of contexts, policy evaluation can be conducted based upon two factors: (1) a policymaker's preferences and (2) the conditional distribution of the outcomes of interest given a policy and available information. What this means is that one of the main objects of interest to scholarly researchers, identification of the true or best model of the economy, is of no intrinsic importance in the policy evaluation context; even though knowledge of this model would, were it available, be very relevant in policy evaluation. Hence model selection, which is a major endeavor in much of empirical macroeconomic research, is not a necessary component of policy evaluation.

In contrast, our argument is that, in many cases, model selection is actually inappropriate, as conditioning policy evaluation on a particular model ignores the role of model uncertainty in the overall uncertainty that exists with respect to the effects of a given policy choice. This is true both in the sense that many statistical analyses of policies do not systematically evaluate the robustness of policies across different model specifications and in the sense that many analyses fail to adequately account for the effects of model selection on statistical inference. In contrast, we advocate the use of model averaging methods, which represent a formal way through which one can avoid policy evaluation that is conditional on a particular economic model.

From the perspective of the theory of policy evaluation, model uncertainty has important implications for the evaluation of policies. This was originally recognized in William Brainard's classic analysis[4], where model uncertainty occurs in the sense that the effects of a policy on a macroeconomic outcome of interest are unknown, but may be described by the distribution of a parameter (one that measures the marginal effect of the policy on the outcome). Much of what we argue in terms of theory may be interpreted as a generalization of Brainard's original framework and associated insights to a broader class of model uncertainty.

An additional advantage of our approach is that it provides a firm foundation for the integration of empirical analysis with policy evaluation. By explicitly casting policy evaluation exercises as the comparison of the losses associated with the distribution of macroeconomic outcomes conditional on alternative policy scenarios, connections between the observed history of the economy and policy advice are seamlessly integrated. Conventional approaches, which often equate evaluation of the efficacy of a policy with the statistical significance of an estimated coefficient, do not embody an equally straightforward way of moving from empirical findings to policy outcomes. Hence, one practical implication of our discussion is that the reporting of empirical results for policy analysis should focus more explicitly on describing probability distributions for outcomes of interest, conditioned on a given policy, rather than on statistical significance testing per se.

Our goals in this paper are ambitious in that we are attempting to place policy theoretical and empirical evaluation exercises in a framework that properly accounts for the decision-theoretic nature of the question and which properly accounts for the different types of uncertainty. As such, we are motivated by similar concerns as have influenced a number of other researchers. Many of James Heckman's contributions may be interpreted as providing methods for policy analysis, specifically policy analysis that properly accounts for the ways in which individuals make decisions.[5] In terms of explicit decision-theory approaches, Gary Chamberlain[6] and Christopher Sims[7] have argued in favor of Bayesian decision-theoretic approaches to data analysis.[8] Charles Manski[9] has, in contexts where one cannot identify which of several models explain a given data set, advocated an approach that focuses on finding undominated policies, i.e. policies that are optimal for at least one model consistent with the data. Our own approach has been strongly influenced by this important work and we will indicate in the course of our discussion where our approach overlaps with and where our approach contrasts with this previous research. And of course, much of what motivates our discussion is modern statistical decision theory, which now functions as a foundation of Bayesian statistics.

We are also far from the first researchers to attempt to integrate concerns about model uncertainty into policy analysis. In terms of general econometric questions, Edward Leamer has made a range of fundamental contributions to the development of

methods of econometric inference that account for model uncertainty.[10] Leamer's ideas have motivated a number of recent developments in the statistics literature.[11] In terms of the theory of policy analysis, Lars Hansen and Thomas Sargent, for example, have pioneered the use of robust control theory to evaluate macroeconomic policy in environments in which model uncertainty may be characterized as occurring around a particular core model.[12] This research program has initiated new directions in policy evaluation which focus on how to construct policies that are robust against unfavorable draws from the space of possible models.

Further, model uncertainty has motivated a range of empirical analyses. The monetary policy rules literature has become quite explicit in this objective. And to be fair, it is rare to see an empirical paper that does not consider some modifications to a given baseline specification to see whether particular empirical claims are robust across modifications.[13] Within the economic growth literature, analyses such as those by Ross Levine and David Renelt[14] and Xavier Sala-i-Martin[15] have modified standard growth regression analysis to account for model uncertainty; Gernot Doppelhofer, Ronald Miller, and Sala-i-Martin and Carmen Fernandez, Eduardo Ley and Mark Steel[16] have explicitly employed the averaging approach to model uncertainty that we endorse.[17] And of course, empirical work very typically involves a consideration of the robustness of findings across different specifications of the estimated model, application of the model to different subsamples of data, etc. It is therefore a caricature of the empirical literature to suggest that model uncertainty is generally ignored. Relative to these applied approaches, we believe our analysis will have some useful suggestions on how to make robustness analyses more systematic and how to link the evaluation of model uncertainty to the goals of an econometric exercise in a more effective fashion.

While our goals are ambitious, it is important to recognize that there are important limits in the extent to which we have achieved them. While the decision-theoretic approach is, in an abstract sense, an extremely appealing way to engage in econometric policy evaluation, there are significant open questions as to how one would implement the approach. We will discuss some ways of making decision theory and model averaging operational, but there is still very substantial work that needs to be done. Finally, we wish to be clear that we do not believe there is "one true path" for empirical

work. Debates concerning the philosophical merits of Bayesian versus frequentist approaches, etc. are of little intrinsic use to us. We are interested in the pragmatic questions that revolve around the use of theoretical and econometric models to inform policy evaluation.

Section 2 of the paper introduces a basic framework for policy evaluation. The discussion in this section is designed to place policy evaluation in a decision-theoretic framework, a framework that we will exploit throughout the paper. Section 3 provides an analysis of how model uncertainty affects policy evaluation. We contrast our perspective with other recent efforts in the economics literature to address model uncertainty. Section 4 explores some theoretical implications of model uncertainty for policy evaluation. Section 5 discusses some issues that arise in implementing the general decision-theoretic framework we have described. First, we show how our basic framework may be applied under Bayesian, frequentist, and Waldean perspectives on policy evaluation. Second, we discuss a number of questions that arise when one is specifying a space of possible models. Section 6 provides two applications of our ideas: monetary policy rules and the analysis of growth policies. These applications are designed to follow previous empirical work closely in order to illustrate how to implement some of the methodological ideas we advocate. Section 7 provides summary and conclusions. Computational and data appendices follow.

2. Decision theory and uncertainty

In this section, we describe a basic decision-theoretic approach to policy evaluation. The abstract ideas we describe constitute the building blocks of modern statistical decision theory.[18] No claim of originality is made. We believe that the underlying logic of the framework is something that the great majority of economists do or would regard as appealing. It is also the case that these ideas have periodically appeared over time as different economists have attempted to place empirical research on a more policy-relevant foundation.[19] Our own discussion will place these ideas in a

context that helps identify some dimensions along which this framework can inform theoretical and empirical work on macroeconomic policy analysis.

From a decision-theoretic perspective, one thinks of a policymaker as facing a choice among a set of outcomes and wishing to use available information, including data on the economy, to inform this choice. As such, the policymaker's decision is interpretable as a standard microeconomic problem of choice under uncertainty. To formalize this idea, suppose that a policymaker must choose a policy, indexed by p from some set of possible policies P. The policymaker has available a data set d (a realization from a process with support D) which may be used to inform the policy evaluation. We initially assume that the policymaker is evaluating policies conditional on a given model of the economy, m. At this level, there is no need to precisely define what constitutes a model; typically a model will incorporate a particular economic theory or theories as well as various functional form specifications. While the model of the economy could be treated as part of the policymaker's information set (which would mean treating it in a symmetric fashion to d), it is convenient to separate it from the other information he possesses. Each policymaker has preferences over policy effects that may be represented as a loss function $l(p,\theta)$ where θ represents whatever quantities affect the function; the support of these unknowns is Θ. For example, θ may represent parameters that determine the effects of the policy. Typically, θ will include innovations to the economy that have not been realized at the time the policy is chosen. From the perspective of a policymaker, uncertainty about θ is the only source of uncertainty about the losses of a given policy. For simplicity, we do not allow the loss function to depend on the model; this generalization may easily be incorporated.

In order to describe the effect of uncertainty over θ on policy evaluation, it is necessary to characterize the policymaker's preferences as they relate to risk. We initially assume that the policymaker is an expected loss minimizer; alternative preference assumptions will be considered later. Expected loss calculations, in turn, require specification of the probabilities associated with different realizations of θ. These probabilities are described by the density $\mu(\theta \mid d,m)$, so that uncertainty about θ

is conditioned on the available data d and a particular model m of the economy. The expected loss associated with policy p is therefore

$$E\left(l\left(p,\theta\right)|d,m\right)=\int_{\Theta}l\left(p,\theta\right)\mu\left(\theta|d,m\right)\mathrm{d}\theta \tag{1}$$

This type of calculation allows for policy comparisons. Optimal policy choice may be treated as

$$\min_{p\in P}\int_{\Theta}l\left(p,\theta\right)\mu\left(\theta|d,m\right)\mathrm{d}\theta \tag{2}$$

As equations (1) and (2) illustrate, policy analysis is thus straightforward once the loss function $l\left(p,\theta\right)$ and the probability density $\mu\left(\theta|d,m\right)$ are specified. However, it is interesting to observe that the sorts of calculations associated with (1) and (2) are not necessarily those that are associated with conventional econometric practice. This is so in three senses.

First, the relevant uncertainty associated with θ cannot necessarily be reduced to its expected value and associated variance. The entire posterior probability density of θ may be relevant. Of course, as has been understood since the early days of mean variance analysis in portfolio theory, there are various assumptions on the structure of uncertainty and policymaker preferences under which the second moments are the only moments of $\mu\left(\theta|d,m\right)$ that affect policy assessment. The appropriateness of these assumptions will differ from context to context and so should not be assumed without any forethought.

Second, even if the relevant uncertainty associated with θ can be summarized by its posterior mean and variance this does not provide a clear way of linking policy evaluation to hypothesis testing. For example, consider the way in which various policies are evaluated in the empirical growth literature. Typically, a researcher identifies an empirical proxy for a policy and determines whether it is relevant for growth according to whether or not it is statistically significant at the 5% level. This assessment does not

7

directly speak to the question of whether the policy variable should be changed, even if one ignores the question of the costs of such a change.

What implications might one draw from these two arguments? One implication is that it is generally more appropriate to report posterior distributions that describe the effects of policies on variables of interest, rather than focus on test statistics per se. The relevance of this implication differs across empirical literatures; the monetary policy rule literature is very much focused on the evaluation of policy rules with respect to loss functions.[20] In contrast, the economic growth literature is very much dominated by hypothesis testing as a way to evaluate growth policies; for example the survey of the empirical growth literature by Robert Barro and Sala-i-Martin[21] typically equates evidence that a policy is relevant for growth with the statistical significance of its associated regression parameter. We will discuss the use of empirical models to evaluate growth policies in more detail in Section 6.

A third criticism of conventional econometric practice concerns the distinction between parameters and estimates of parameters The uncertainty that is relevant for policy evaluation is uncertainty over θ, not uncertainty with respect to estimates of the parameter, i.e. $\hat{\theta}$. Yet most empirical work reports standard errors of estimates rather than uncertainty concerning underlying parameters. This is a standard objection Bayesians make of frequentist approaches to econometrics.[22] The import of this criticism will differ across contexts. The reason for this is that for a large range of cases Bayesian and maximum likelihood estimates converge, so that the distinction focusing on the distribution of parameters versus associated estimates is of second-order importance in large samples.[23] We will not focus on this issue further.

3. Model uncertainty

The basic framework we have described may be employed to understand how to account for model uncertainty. To see how one would do this, suppose that there exists a set M of possible models of the economy. We treat the set of possible models as finite; allowing for richer model spaces may be done in a straightforward fashion for a number

of contexts.[24] With respect to our previous discussion, the question we now address is how to incorporate uncertainty about the appropriate model of the economy when evaluating policies.

One important issue in dealing with model uncertainty concerns whether it should be treated in the same way as uncertainty over other unknowns, e.g. parameters or over the realizations of future shocks to the economy. For now, we treat all uncertainty symmetrically, so that the incorporation of model uncertainty into policy evaluation calculations requires only that the policymaker incorporate a probabilistic description of model uncertainty into (1) and (2); however, there will turn out to be some dimensions along which model uncertainty may warrant a different treatment.

i. Expected loss calculations under model uncertainty

In order to extend our discussion in Section 2 to include model uncertainty, it is necessary to modify the description of uncertainty over θ in such a way that it no longer is conditioned on a given model. Put differently, from the perspective of policy evaluation, a policymaker will not want to condition decisions on a particular model unless one knows that the model is true with probability 1. Rather, he will want to compute expected losses conditioning only on the realized data d. Relative to the expected loss calculation described by (1), accounting for model uncertainty means that the expected loss to a policy should be evaluated under the assumption that the model m is an unknown. This means that when the true model m is unknown, the policy evaluation equation (1) should be modified so that the expected loss associated with each policy accounts for this; the expected loss associated with a policy that only conditions on the data may be calculated as

$$E\big(l(p,\theta)\big|d\big) = \int_{\Theta} l(p,\theta)\mu\big(\theta\big|d\big)\mathrm{d}\theta \tag{3}$$

where

$$\mu(\theta|d) = \sum_{m \in M} \mu(\theta|d,m)\mu(m|d) \qquad (4)$$

The term $\mu(\theta|d)$ describes the posterior probability of the relevant unknowns conditional on the observed data d and accounting for model uncertainty. As before, the role of econometric analysis is in computing this object.

Eq. (4) illustrates how one can eliminate dependence of expected loss calculations on a particular model: one treats the identity of the "true" model as an unobserved random variable and "integrates" it out of the loss function and the posterior density for unobservables. This technique is known as model averaging in the statistics literature.[25]

The failure to account systematically for model uncertainty is, in our judgment, a defect of much current econometric practice. "Standard" econometric practice consists of calculating quantities that are variants of the conditional probability $\mu(\theta|d,m)$. As we have argued, in the presence of model uncertainty, the natural object of interest in policy evaluation is $\mu(\theta|d)$. While it is common practice to evaluate the robustness of $\mu(\theta|d,m)$ relative to some set of modifications of a baseline model specification, these are typically ad hoc. In addition, the common practice of reporting results for a set of related models in order to show the robustness or nonrobustness of a given finding across models does not provide a way of combining information across specifications. Nor does this practice provide a clear way of thinking about nonrobustness. If a coefficient is large in one regression and small in another, what conclusion should be drawn? The calculation of $\mu(\theta|d)$ renders such questions moot, as the information about θ that is contained in each model specification is integrated into its construction.

In order to understand what is needed to construct $\mu(m|d)$, it is useful to rewrite this conditional probability as

$$\mu(m|d) = \frac{\mu(d|m)\mu(m)}{\mu(d)} \propto \mu(d|m)\mu(m) \qquad (5)$$

where "$\propto$" means "is proportional to." As eq. (5) indicates, the calculation of posterior model probabilities depends on two terms. The first term, $\mu(d|m)$, is the probability of the data given a model, and so corresponds to a model-specific likelihood. The second term, $\mu(m)$, is the prior probability assigned to model m. Hence, computing posterior model probabilities requires specifying prior beliefs on the probabilities of the elements of the model space M. The choice of prior probabilities for a model space is an interesting and not fully understood problem and will be discussed below. One common choice for prior model probabilities is to assume that each model is equally likely. But even in this case, the posterior probabilities will not be equal since these probabilities depend on the relative likelihoods of each model.

One can develop some insight into what this approach can accomplish by comparing it to the analysis by Andrew Levin and John Williams which is very much in the spirit of model averaging.[26] In their paper, monetary policy rules are evaluated when a forwards-looking model, a backwards-looking model and a forwards/backwards looking hybrid model of output and inflation are each given a probability weight of 1/3; in each case the parameters are also assumed known a priori. The calculation of expected losses from policy rules is done using their analog to eq. (3). Relative to this approach, we would argue that the appropriate model weights are not fixed probabilities but rather posterior probabilities that reflect the relative goodness of fit across the various models. In addition, we would argue that one needs to account for specification uncertainty for each of the models Levin and Williams consider. For example, one would not want to assume lag lengths are known a priori. In other words, model uncertainty occurs at a range of levels including both the economic theory that constitutes the underlying logic of a model as well as the detailed specification of its statistical structure. (Our approach would also account for parameter uncertainty in the calculation of expected losses, but this is a distinct issue from model uncertainty.)

How does model uncertainty alter the ways in which one thinks about statistical quantities? Suppose that the goal of an exercise is to characterize aspects of an unknown quantity δ. Suppose that one is able to calculate the mean and variance of this object conditional on a given model. In order to compute the mean and variance of δ without

conditioning on a given model, one uses the posterior model probabilities to eliminate this dependence. Following formulas due to Leamer,[27] the mean and variance of δ, once one has accounted for model uncertainty are

$$E(\delta|d) = \sum_{m \in M} \mu(m|d) E(\delta|d,m) \tag{6}$$

and

$$var(\delta|d) = E(\delta^2|d) - \left(E(\delta|d)\right)^2 =$$
$$\sum_{m \in M} \mu(m|d)\left(var(\delta|d,m) + \left(E(\delta|d,m)\right)^2\right) - \left(E(\delta|d)\right)^2 = \tag{7}$$
$$\sum_{m \in M} \mu(m|d) var(\delta|d,m) + \sum_{m \in M} \mu(m|d)\left(E(\delta|d,m) - E(\delta|d)\right)^2$$

respectively.

These formulas illustrate how model uncertainty affects a given parameter estimate. First, the posterior mean of the parameter is a weighted average of the posterior means across each model. Second, the posterior variance is the sum of two terms. The first term, $\sum_{m \in M} \mu(m|d) var(\delta|d,m)$, is a weighted average of the variances for each model and directly parallels the construction of the posterior mean. The second term reflects the variance across models of the expected value for δ; these differences reflect the fact that the models are themselves different. This term, $\sum_{m \in M} \mu(m|d)\left(E(\delta|d,m) - E(\delta|d)\right)^2$, is not determined by the model-specific variance calculations and in this sense is new, capturing how model uncertainty increases the variance associated with a parameter estimate relative to conventional calculations. The term measures the contribution to the variance of δ that occurs because different models produce different estimates $E(\delta|d,m)$. To see why this second term is interesting, suppose that $var(\delta|d,m)$ is constant across models. Should one conclude that the overall variance is equal to this same value? In general, one should not do so. So long as there is any variation in $E(\delta|d,m)$ across models, then $var(\delta|d,m) < var(\delta|d)$; the cross model variations in the mean increase the uncertainty (as measured by the variance) that exists with respect to

δ . As argued by David Draper,[28] this second term explains why one often finds that the predictions of the effect of a policy often grossly underestimate the actual uncertainty associated with the effect.

ii. Model uncertainty and ambiguity aversion

This analysis of model uncertainty may be generalized to allow for preferences that move beyond the expected utility paradigm that underlies equations such as (1). In particular, the framework may be adapted to allow for preference structures that evaluate uncertainty about models differently from other types of uncertainty.   Does this distinction between sources matter? We would argue that this is an important implication of some of the work associated with both the new behavioral economics[29] and with recent developments in economic theory.

One famous example of a behavioral regularity that suggests that individual preferences cannot be modeled using standard expected utility formulations is the Ellsberg paradox,[30] which is based on the following experiment. Individuals are asked to state their preferences across 4 different lotteries. For lottery 1, the agent receives a prize of $\$a$ if a red ball is drawn from an urn with 50 red and 50 black balls; lottery 2 is a payment of $\$a$ if a black ball is drawn from the same urn.   Lottery 3 is a payment of $\$a$ for a red ball from a second urn, but where the number of red and black balls is not specified.   Lottery 4 is a payment of $\$a$ for a black ball from urn 2. Daniel Ellsberg argues that individuals show a consistent preference for lotteries 1 and 2 over either 3 or 4. From the perspective of expected utility theory, this is paradoxical as it implies certain violations of the Savage axioms that underlie expected utility theory. For our purposes, the Ellsberg paradox is interesting because it suggests a distaste for model uncertainty, in the sense that lotteries 3 and 4 are associated with a range of possible red/black ball probabilities.

A range of experimental studies have confirmed that individual preferences reflect a distaste for model uncertainty of the type Ellsberg described.[31]   This distaste does not appear to be explained either by the possibility that participants in these experiments do not understand the rules of conditional probability: Paul Slovic and Amos Tversky[32]

found that providing participants with written explanations of why preferring lotteries 1 and 2 to 3 and 4 is inconsistent with expected payoff maximization does not eliminate the paradox. Further, it does not appear that the distaste for urns with unknown proportions reflects a belief that lotteries 3 and 4 are somehow rigged against the participant (in the sense, for example, that the composition of the second urn is changed once a payoff rule is chosen).[33] It therefore seems that the Ellsberg paradox reflects something about individual preferences, not cognitive limitations.

This type of behavior has been axiomatized in recent work by Larry Epstein and Tau Wang and Itzhak Gilboa and David Schmeidler on ambiguity aversion.[34] This work has proposed a reformulation of individual preferences so that they reflect a dislike of "ambiguity" as well as risk. In these approaches, distaste for ambiguity means that the actor places extra weight on the worst uncertain outcome that is possible in a given context. The theoretical development of models of ambiguity aversion is important in showing that ambiguity aversion emerges as a feature of behavior not because of cognitive limitations of an actor but rather from a particular formulation of how an actor evaluates uncertainty in outcomes.

The ideas that underlie recent work on ambiguity aversion are directly applicable to the formulation of policymaker preferences. Notice that one essential feature in the lotteries that motivate the Ellsberg paradox appears to be the distinction agents draw between knowing that an urn has 50 red and 50 black balls versus not knowing the proportions of colors, even if one is then allowed to choose which color produces a payoff. This is interpretable as meaning that individuals assess model uncertainty differently from uncertainty with respects to outcomes within a model. While urn experiments of course do not directly measure objects that are relevant to policymaker preferences, we do believe they suggest that model uncertainty plays a special role in such preferences.

In our context, suppose that a policymaker's preferences reflect ambiguity aversion in the sense that extra weight is placed on the most unfavorable model of the economy that may hold, relative to the weight associated with the posterior probability of that model. Following the approach suggested by Epstein and Wang[35], such preferences may be formalized through the function

$$(1-e)\int_{\Theta} l\left(p,\theta\right)\mu\left(\theta|d\right)\mathrm{d}\theta + e\left(\sup_{m\in M}\int_{\Theta} l\left(p,\theta\right)\mu\left(\theta|d,m\right)\mathrm{d}\theta\right) \tag{8}$$

In this expression, e indexes the degree of ambiguity aversion. When $e=0$, this expression reduces to our earlier expected loss calculation. When $e=1$, policies are evaluated by a minimax criterion: the loss associated with a policy is determined by the expected loss it produces under the worst possible model; good rules are those that minimize losses under worst case scenarios.[36]

Is this type of preference structure relevant to policy analysis? We argue that it is on several levels. First, the preference structure does reflect the sorts of experimental evidence that has motivated the new behavioral economics, and so as a positive matter may be useful in understanding policymaker preferences. Second, we believe this type of preference structure reflects the intuition that there exist qualitative differences across types of uncertainty. In particular, we believe that ambiguity aversion is a way of acknowledging that one can plausibly argue that there are situations where priors over the space of models are not necessarily well enough defined, nor is any version of a noninformative prior well enough developed, so that standard expected loss calculations can be sensibly made. And of course, as work by Epstein and Wang and Gilboa and Schmeidler has shown, ambiguity aversion is perfectly consistent with rational decisionmaking; the expected utility paradigm does not have a privileged position in this sense.

iii. Relation to other work

The approach we advocate to incorporating model uncertainty may be usefully contrasted with a number of research programs.

Extreme bounds analysis

An important research program on model uncertainty originates with Edward Leamer and includes a strategy for rendering the reporting of empirical results more

credible. Leamer's ideas have been most extensively developed in the context of linear regressions. Suppose that one is interested in the relationship between an outcome y and some variable p. There exists a set Z of other variables that may or may not affect y as well. For each subset of regressors Z_m (different subsets of Z correspond to different models), one can evaluate the effect of p on y via the regression

$$y_i = \delta_m p_i + \beta'_m Z_{m,i} + \varepsilon_i \tag{9}$$

Leamer proposes evaluating evidence on the relationship between p and y via the distribution of estimates $\hat{\delta}_m$ across different subsets of control variables. He argues that a benchmark for evaluating the robustness of such inferences is the stability of the sign of $\hat{\delta}_m$ across different specifications. Leamer proposes a rule of thumb which stipulates that the relationship between x and y should be regarded as fragile if the sign of $\hat{\delta}_m$ changes across specifications.

Following work by William Brock and Steven Durlauf,[37] this rule of thumb may be given a decision-theoretic interpretation. Suppose that a policymaker is considering whether to change p from an initial value $\bar{p}$ to some alternative $\bar{\bar{p}} > \bar{p}$ Suppose that conditional on model m, the loss function for the policymaker is $-\hat{\delta}_m\left(\bar{\bar{p}} - \bar{p}\right)$. Leamer's rule means that one will choose to implement the policy if and only if $\inf_{m \in M} \hat{\delta}_m \left(\bar{\bar{p}} - \bar{p}\right) > 0$. This illustrates how in two respects Leamer's rule presupposes rather special preferences on the part of the analyst. First, the rule requires that $\hat{\delta}_m$ is a sufficient statistic for the policymaker's payoff function conditional on a particular model. Second, the rule means that the policymaker's evaluation of risk is described by a very particular functional form.

Extreme bounds analysis has been subjected to serious criticisms by a number of authors.[38] The major criticism of the method, in our reading of the literature, has been that Leamer's procedure is insensitive to the relative goodness of fit of different models. We believe this concern is valid: the fact that a model that appears to be grossly misspecified produces a different sign for $\hat{\delta}_m$ than is found in a model that does not

16

appear to be misspecified, intuitively seems a weak reason to conclude that evidence concerning δ is fragile. This does not, however, mean that Leamer's deep idea that one needs to account for the fragility of regression findings across specifications is invalid or that extreme bounds analysis cannot be adapted in a way to respond to the objection.

Following an argument by Brock and Durlauf,[39] one can modify Leamer's idea in a way that preserves its core intuition. This becomes apparent when one interprets Leamer's analysis in the context of the ambiguity aversion analysis we described above. Specifically, the decision-theoretic version of extreme bounds analysis is a limiting case of eq. (8) above where $e = 1$ and $\int_{\Theta} l(p,\theta) \mu(\theta|d,m) \mathrm{d}\theta = -\hat{\delta}_m p$. This calculation makes clear that ambiguity aversion is the key feature underlying extreme bounds analysis as a procedure for reporting empirical results. This implies that if one relaxes the requirement that $e = 1$, one can preserve the ambiguity aversion that lies at the core of the extreme bounds method and at the same time address criticisms of the procedure. In particular, for $0 < e < 1$, the overall effect of a particular model-specific parameter on a policy evaluation will be increasing in the model's posterior probability.

Robust optimal control

In an influential recent body of research, Lars Hansen and Thomas Sargent have employed robust decision theory to account for the fact that a policymaker typically does not know the true model of the economy.[40] This work has stimulated a growing literature.[41] The robust control framework differs from ours in two respects. First, Hansen and Sargent consider model uncertainty that is centered around a "core model." What this means is that they consider environments in which the true model is known only up to some local neighborhood of models that surround the core model. This neighborhood set may be small or quite large depending on how the notion of distance between models is parameterized. We will call this type of analysis a local analysis even though technically speaking the neighborhood does not have to be small in the usual mathematical sense.

Second, Hansen and Sargent do not work with priors on the model space, i.e. $\mu(m)$. Rather, they engage in minimax analysis, in which the least favorable model in the space of potential models is assumed to be the "true" one for purposes of policy evaluation; this assumption is in the spirit of Abraham Wald.[42] To put it another way, Hansen and Sargent assume that Nature draws a model from the neighborhood set of models in order to maximize cost to the policymaker. They then set their policy rule in order to minimize cost while playing such a game against Nature. In fact, their analysis is explicitly based on a two player zero sum game where Nature chooses a model (from a set of models centered at a core model) in order to maximize losses to the policymaker and the policymaker chooses a policy to minimize losses.

Our discussion of the decision-theoretic approach to policy analysis is closely connected to the Hansen-Sargent research program. In comparison to our discussion, Hansen and Sargent may be interpreted as developing their analysis on the basis of a particular way of characterizing the space of potential models (one that possesses enormous power because it allows one to bring to bear robust control theory tools) combined with a description of policymaker preferences in which $e = 1$. This approach reflects a modeling philosophy in which one starts with a well-developed and economically sensible core model and explores the implications of allowing for the possibility that the core model is misspecified. As Hansen and Sargent describe their approach: [43]

> Starting from a single dynamic model, we add perturbations that represent potential model misspecifications around that benchmark model. The perturbations can be viewed as indexing a large family of dynamic models…We prefer to think about the perturbations as errors in a convenient, but misspecified, dynamic macroeconomic model. We take the formal structure for perturbations from a source that served macroeconomists well before…

Our analysis is motivated by the belief that model uncertainty is, in many macroeconomic contexts, associated with the existence of more than one core model that potentially describes the phenomenon under study. Disagreements as to whether democratization is necessary for sustained growth or whether business cycles are better understood as generated by monetary versus real factors are associated with very

different conceptions of the macroeconomy and constitute a different type of uncertainty from the sort for which robust control theory is best designed. Hence, we favor an approach that allows for model uncertainty across a range of core models.[44] As such, it attempts to address the sorts of challenges John Taylor has identified in modern research on monetary policy: [45]

> …researchers are using many different types of models for evaluating monetary rules, including small estimated or calibrated models with or without rational expectations, optimizing models with representative agents, and large econometric models with rational expectations. Some models are closed economy models, some are open economy models, and some are multicountry models…Seeking robustness of…rules across a wide range of models, viewpoints, historical periods, and countries, is itself an important objective of policy evaluation research.

Our focus on "global" (in this sense) model uncertainty has implications for how one thinks about losses. Specifically, if one does not believe that the space of potential models is "narrow" in the sense defined by Hansen and Sargent, the minimax approach is likely to give highly unsatisfactory results. The reason is that the minimax assumption implies that policy evaluation will ignore posterior model probabilities. Hence a model with arbitrarily low posterior probability can determine the optimal policy so as long it represents the "worst case" in terms of loss calculations. This does not mean that the minimax assumption in Hansen and Sargent is in any sense incorrect, only that the appropriateness of a particular strategy for evaluating losses depends on context. In particular, we believe that the minimax strategy is very natural for the study of local forms of model uncertainty that are explored in the new robust control approach to macroeconomics. In fact, the minimax approach has proven extremely important in the development of robust approaches to policy evaluation, which is arguably the main new theoretical contribution of recent macroeconomic research on model uncertainty. In the next section, we show how a very localized version of the minimax strategy can be developed that gives intuitively reasonable results and uses only simple calculus tools.

4. Theoretical implications

In this section, we consider some theoretical implications of model uncertainty for policy evaluation. Specifically, we analyze how a preference for policy robustness influences the design of policies. This approach employs minimax preferences in the context of analyzing how a policymaker might account for the introduction of model uncertainty defined by a local neighborhood of models generated around a benchmark model or set of models. As we have suggested, robustness analysis represents an important innovation in the theory of policy evaluation and may be interpreted as an approach to accounting for model uncertainty when policymaker preferences reflect ambiguity aversion.[46]

i. local robustness analysis

We first describe an approach to conducting local robustness exercises in policy design. To do this, recall that the general discussion in Sections 2 and 3 placed primary focus on the role of the posterior density of θ, $\mu(\theta|d,m)$ if the model is known, $\mu(\theta|d)$ if the model is unknown, in allowing a policymaker to evaluate policies. We will initially assume that m is known and ask how perturbations around this initial model affect optimal policy choice. Specifically, we will ask how the optimal policy changes with respect to a change in one of the parameters of that density, which we designate as α. Let $p^*(\alpha)$ denote the optimal policy as a function of this parameter and let $J(p^*(\alpha),\alpha|m)$ denote the value of (2) evaluated at this optimal policy choice. For technical simplicity, we assume that both $J(p^*(\alpha),\alpha|m)$ and $p^*(\alpha)$ are both twice differentiable.

To think about robustness we consider how a policy should be chosen when the policymaker does not choose it in response to a fixed parameter $\bar{\alpha}$ but rather chooses it when the parameter is constrained to lie in a neighborhood $N = [\bar{\alpha} - \Delta, \bar{\alpha} + \Delta]$. Each element in this neighborhood defines a different distribution for θ and thus constitutes a

separate model. Of course, one cannot specify an optimal policy unless one specifies how this parameter is determined. The key idea behind robustness analysis is to assume that this choice is dictated in a way that is least favorable to the policymaker. Metaphorically, one can suppose that the policymaker faces an "adversarial agent" (AA) who chooses the actual parameter from this interval in order to maximize the loss function of the policymaker. This metaphor captures the idea in robustness analysis that one chooses a policy based upon minimax considerations. A robust policy is one that is optimal against the least favorable model in the space of models implied by the neighborhood.

To understand how robustness affects optimal policy choice, we first consider how an adversarial agent will choose an element of N. When Δ is small, one can work with the approximation

$$J\left(p^*\left(\bar{\alpha}+\Delta\right),\bar{\alpha}+\Delta\big|m\right)$$

$$\approx J\left(p^*\left(\bar{\alpha}\right),\bar{\alpha}\big|m\right)+\left(\frac{\partial J\left(p^*\left(\bar{\alpha}\right),\bar{\alpha}\big|m\right)}{\partial p}\frac{\partial p^*\left(\bar{\alpha}\right)}{\partial \alpha}+\frac{\partial J\left(p^*\left(\bar{\alpha}\right),\bar{\alpha}\big|m\right)}{\partial \alpha}\right)\Delta \qquad (10)$$

$$\approx J\left(p^*\left(\bar{\alpha}\right),\bar{\alpha}\big|m\right)+\frac{\partial J\left(p^*\left(\bar{\alpha}\right),\bar{\alpha}\big|m\right)}{\partial \alpha}\Delta$$

The second equality follows from the envelope theorem. Hence, the adversarial agent will, for small Δ, choose $\bar{\alpha}+\Delta$ if $\dfrac{\partial J\left(p^*\left(\bar{\alpha}\right),\bar{\alpha}\big|m\right)}{\partial \alpha}>0$, $\bar{\alpha}-\Delta$ otherwise.

The robust policy response can thus be computed as a response to the action of the AA. It is straightforward to show that the robust policy response to the introduction of the AA is[47]

$$\mathrm{d}p^* = -\left(\frac{\partial^2 J\left(p^*(\bar{\alpha}),\bar{\alpha}\,|m\right)}{\partial p \partial \alpha}\middle/\frac{\partial^2 J\left(p^*(\bar{\alpha}),\bar{\alpha}\,|m\right)}{\partial p^2}\right)\Delta \quad \text{if} \quad \frac{\partial J\left(p^*(\bar{\alpha}),\bar{\alpha}\,|m\right)}{\partial \alpha} > 0$$

$$(11)$$

$$\mathrm{d}p^* = \left(\frac{\partial^2 J\left(p^*(\bar{\alpha}),\bar{\alpha}\,|m\right)}{\partial p \partial \alpha}\middle/\frac{\partial^2 J\left(p^*(\bar{\alpha}),\bar{\alpha}\,|m\right)}{\partial p^2}\right)\Delta \quad \text{if} \quad \frac{\partial J\left(p^*(\bar{\alpha}),\bar{\alpha}\,|m\right)}{\partial \alpha} < 0;$$

One important feature of these formulas is that they indicate how introducing an adversarial agent and considering robustness is different from simply introducing uncertainty around a model parameter. As first shown in the classic work of Kenneth Arrow and John Pratt, risk is a second order phenomenon in the sense that starting, for example, with consumption at a certain risk free level, the addition of a sufficiently small mean zero random variable to this consumption level has no effect on utility. In our context, adding a small amount of uncertainty around $\bar{\alpha}$ in the form of a zero mean random variable would similarly have no effect on optimal policy. The introduction of a neighborhood of uncertainty around $\bar{\alpha}$ combined with an adversarial agent, in contrast, produces a first order effect on behavior, except for the special case $\frac{\partial J\left(p^*(\bar{\alpha}),\bar{\alpha}\,|m\right)}{\partial \alpha} = 0$. The reason is quite intuitive: the presence of the adversarial agent ensures the effect on the expected loss to the policymaker from the introduction of the neighborhood will never be zero. Put differently, robustness analysis is predicated on the idea that uncertainty cannot be modeled as a mean preserving spread, but rather is measured in terms of the bounds of the effects of the uncertainty on changes in payoffs. For this reason, robustness analysis is conceptually distinct from conventional risk analysis.

application to Brainard

This general discussion can be applied in the context of Brainard's classic analysis of optimal choice of policies in the presence of uncertainty.[48] Brainard's model focuses on the question of how to stabilize (in the sense of minimizing expected squared

deviations) a variable y around some objective $\bar{y}$ using two policy instruments p_1 and p_2. The baseline model for this analysis is

$$y = \theta_1 p_1 + \theta_2 p_2 + \varepsilon \tag{12}$$

where ε denotes a random variable that captures aspects of y outside the policymaker's influence. In the context of our loss framework, Brainard's problem may be written as

$$\min_{(p_1,p_2)} \int_{\Theta} (\theta_1 p_1 + \theta_2 p_2 + \varepsilon - \bar{y})^2 \mu(\theta_1, \theta_2 | d, m) d\theta_1 d\theta_2 \tag{13}$$

Following Brainard, it is assumed that ε is independent of θ_1 and θ_2 and that $E(\theta_1) = E(\theta_2) = 1$. Letting σ_{ij} denote the covariance of θ_i and θ_j, Brainard shows that the optimal policy choices in this environment are

$$p_1^* = \frac{\sigma_{22} - \sigma_{12}}{(\sigma_{11}+1)(\sigma_{22}+1) - (\sigma_{12}+1)^2} (\bar{y} - E(\varepsilon)) \tag{14}$$

and

$$p_2^* = \frac{\sigma_{11} - \sigma_{12}}{(\sigma_{11}+1)(\sigma_{22}+1) - (\sigma_{12}+1)^2} (\bar{y} - E(\varepsilon)) \tag{15}$$

The key insight of these formulas is that policy choices with uncertain effects as formulated here render the choice of policies analogous to a portfolio problem such that the policy weights are determined by an optimal mean/variance tradeoff.

How does a robustness analysis affect these calculations? In order to do this we consider how, starting from fixed parameters, allowing for an adversarial agent to choose a parameter from an interval centered on these parameters affects optimal policy. Let $\bar{\sigma}_{ij}$ denote the baseline for parameter σ_{ij}. Suppose that the adversarial agent chooses the

variance of the first instrument from the interval $[\bar{\sigma}_{11} - \Delta, \bar{\sigma}_{11} + \Delta]$. Using (13), it is straightforward to verify that the AA will choose $\bar{\sigma}_{11} + \Delta$, since the policymaker's payoff is decreasing in the variance of the policy instrument's parameter, i.e. the loss is increasing in σ_{11}. (This follows immediately from the risk aversion built into the policymaker's loss function.) The first order conditions for optimal policy choice may be shown to imply

$$\frac{\mathrm{d}p_1^*\left(\bar{\sigma}_{11}\right)}{\mathrm{d}\sigma_{11}} = -\frac{p_1^*\left(1+\bar{\sigma}_{22}\right)}{\left(1+\bar{\sigma}_{11}\right)\left(1+\bar{\sigma}_{22}\right)-\left(1+\bar{\sigma}_{12}\right)^2} \tag{16}$$

and

$$\frac{\mathrm{d}p_2^*\left(\bar{\sigma}_{11}\right)}{\mathrm{d}\sigma_{11}} = -\frac{\mathrm{d}p_1^*\left(\bar{\sigma}_{11}\right)}{\mathrm{d}\sigma_{11}}\frac{1+\bar{\sigma}_{12}}{1+\bar{\sigma}_{22}} \tag{17}$$

Equations (16) and (17) illustrate several basic ideas. First, policy 2 is always adjusted in the opposite direction to policy 1 if $1+\bar{\sigma}_{12} > 0$ and in the same direction if $1+\bar{\sigma}_{12} < 0$. Recall that the policies have been normalized so that the expected values of their effects are 1, i.e. θ_i has been divided by $E(\theta_i)$. This suggests a presumption that the policies will be adjusted in opposite directions.

Second, regardless of the covariance structure of the policy effects, an increase in σ_{11} leads to a reduction in $\left|p_1^*\right|$. This makes intuitive sense: the less trustworthy control is used less aggressively. Combined with $1+\bar{\sigma}_{12} > 0$, one has a "precautionary principle" for policymakers: one robustifies against uncertainty in policy 1 by using that policy less aggressively and policy 2 more aggressively.

Third, this discussion illustrates the difference between evaluating the introduction of risk and robustness. Suppose that one started with $\sigma_{11} = \sigma_{22} = 0$ and began a local increase in the variances. Following the logic of the Arrow-Pratt theory of

risk aversion, there would not be a first order effect. The robustness analysis, in contrast, does produce a first order effect.

application to monetary policy rule evaluation

Similar results apply to the question of monetary policy rules. This can be seen using a model by Lars Svensson[49] which represents a one equation version of an important output/inflation model due to Glenn Rudebusch and Svensson.[50] In this model, π_t denotes the gap between actual inflation and some target, y_t denotes the gap between output and some target, and e_t denotes an i.i.d. sequence of shocks. The inflation gap evolves according to

$$\pi_{t+1} = \phi \pi_t + \delta y_t + e_t \tag{18}$$

where $\phi = 1$. This equation is a proxy for the actual policy process, i.e. here the policymaker is assumed to be able to control the output gap. The policymaker's preferences are described by the loss function

$$L = E_t \left(\sum_{j=0}^{\infty} \beta^j (\pi_{t+j}^2 + \lambda y_{t+j}^2) \right) \tag{19}$$

Svensson shows that the optimal policy rule for this model is

$$y_t^* = -\frac{\beta \delta k}{\lambda + \beta \delta^2 k} \pi_t \tag{20}$$

where $k = \dfrac{1}{2} \left(1 - \dfrac{\lambda(1-\beta)}{\beta \delta^2} + \sqrt{1 + \dfrac{\lambda(1-\beta)}{\beta \delta^2} + \dfrac{4\lambda}{\delta^2}} \right)$.

To see how robustness works for this model, consider the coefficient ϕ, in (18), which is assumed to equal 1 in all periods by Svensson. Suppose that at time t the adversarial agent may select ϕ from the neighborhood $N = [1 - \Delta, 1 + \Delta]$ for period t;

there is no such choice for future periods. One can show that the loss to the policymaker is increasing in this coefficient, so the least favorable possible coefficient in N is $1+\Delta$. (Intuitively, a policymaker prefers less persistence in the inflation process as it diminishes the net costs to an expansionary policy today.) The optimal choice of the output gap will then equal

$$y_t^* = -(1+\Delta)\frac{\beta\delta k}{\lambda+\beta\delta^2 k}\pi_t \qquad (21)$$

which is more aggressive than the original rule. To understand the difference, robustness in this case means that the policymaker needs to react more aggressively when inflation experiences a shock due to the potentially explosive dynamics associated with the least favorable coefficient $\phi = 1+\Delta$. The locally robust response to this potential for explosiveness in the inflation process is to act more aggressively in response to deviations of output above target. This finding is consistent with the intuition when the channel from the control variable to the outcome of interest is more "trustworthy" than the other determinants of the outcome of interest (the free dynamics of the process) in the sense that if one robustifies with respect to those parameters that characterize the free dynamics, one will use the control more aggressively.[51]

Alternatively, robustness may be sought with respect to the measure of control strength δ, i.e. rather than treat the control strength as a fixed $\bar{\delta}$, the measure of control strength is chosen from the neighborhood $[\bar{\delta}-\Delta, \bar{\delta}+\Delta]$ by an adversarial agent. One can show that the least favorable parameter for the policymaker in this neighborhood is $\bar{\delta}-\Delta$. This is unsurprising as a smaller value for δ in (18) implies a steeper Phillips curve for the policymaker. The response to this change will depend on the sign of $\beta k\delta^2 - \lambda$. If this term is positive, then the policymaker will be more aggressive than occurs when there is no desire to make policies robust with respect to δ. In other words, the coefficient that relates π_t to y_t^* will be larger than appears in (20). On the other hand, if this term is negative, the coefficient will be smaller than appears in (20). Why does the effect of introducing robustness affect policy responses in this way? The condition $\beta k\delta^2 - \lambda > 0$ implies that relatively little weight is placed upon output gap

26

volatility. This leads the policymaker to react very strongly when output is above target; a central bank with such preferences can choose a robust policy strategy to guard against uncertain control by becoming more aggressive in moving output back down to target.

It is interesting to compare this result with the following statement by John Taylor[52]

> I think it is clear that the Phillips curve and the low estimate of the natural rate of unemployment led to the appointment of policymakers with less concern about pursuing price stability. It also probably led to monetary decisions – such as delays in raising interest rates when faced with inflationary pressures in the late 1960's and 1970's – which were inconsistent with price stability.

Suppose one interpreted Taylor as saying that policymakers in the late 1960's and early 1970's had high confidence in their Phillips curve slope estimates, i.e. δ was close to zero. As confidence waned and Δ became larger during the experience of the stagflation in the 1970's, our findings suggest that control would have become more aggressive so long as $\beta k \delta^2 - \lambda > 0$, which would be consistent with the preferences of an inflation "hawk" such as Paul Volcker or Alan Greenspan.

Of course, we do not claim that such a simple model can explain the US monetary history over the last 25 years. We only offer this scenario to illustrate how robustness analysis can yield interpretable results. More generally, we believe that robustness analysis is important in allowing one to analyze how "ignorance" affects policy, where ignorance is measured using the intervals around parameters.

ii. robustness with multiple core models

The analysis of robustness may be extended to the case where there is more than one core model. Abstractly, the analysis of robustness with respect to a parameter α of $\mu(\theta|d)$ may still be done using formula (11) if $J\left(p^*(\alpha), \alpha|m\right)$ is replaced with $J\left(p^*(\alpha), \alpha\right)$ where

$$J\left(p^{*}(\alpha),\alpha\right) = \min_{p \in P} \int_{\Theta} l(p,\theta) \left(\sum_{m \in M} \mu(\theta|d,m)\mu(m|d) \right) d\theta \qquad (22)$$

so that $p^{*}(\alpha)$ now denotes the optimal policy conditional on α after model uncertainty has been accounted for.

We will use (22) as the basis for our discussion of robustness with multiple core models. In doing so, we will not address issues of robustness that arise when ambiguity aversion is present in the form described by (8), although one can certainly conduct our analysis under such preferences.

application to growth economics

In order to see what new insights emerge when one introduces multiple core models, we develop a robustness analysis in a growth context. First, we discuss within model robustness and then allow for multiple core models.

Consider a policymaker who is evaluating whether to change a policy variable p in order to affect a given country's growth rate. We consider the econometric issues involved with such a question below; here we wish to deal with some theoretical issues. Let model m of the growth process equal

$$g_i = v_m' S_{i,m} + \delta_m p_i + \varepsilon_i = \delta_m p_i + \upsilon_{im} \qquad (23)$$

Here, $S_{i,m}$ denotes all growth determinants other than the policy variable p_i; different models are indexed by different choices of growth determinants. Suppose this regression is applied to data in order to produce estimates of the mean and variance of δ as well as the covariance of δ and υ.

Let the policymaker evaluate growth in country i according to the loss function

$$-Eg_i + \frac{r}{2}\sigma_{g_i g_i} \qquad (24)$$

The optimal policy level for a given model will, under these preferences, equal

$$p_i^* = \frac{E(\delta_m) - r\sigma_{\delta \upsilon_i, m}}{r\sigma_{\delta \delta, m}} \tag{25}$$

How does one design a robust policy strategy to deal with uncertainty in the effectiveness of policy parameter $\sigma_{\delta\delta}$? Taking $\bar{\sigma}_{\delta\delta}$ as the value of the parameter without uncertainty, following the same line of argumentation used above, the policymaker does this by choosing a policy that guards against the least favorable value in the interval $[\bar{\sigma}_{\delta\delta} - \Delta, \bar{\sigma}_{\delta\delta} + \Delta]$. The least favorable value is $\bar{\sigma}_{\delta\delta} + \Delta$, since the policymaker is assumed to be risk averse. In turn, the optimal policy choice has the property that

$$dp^* = -\frac{p^*}{\sigma_{\delta\delta}}\Delta \tag{26}$$

which means that the robust policy level $p^* + dp^*$ is smaller than $p^*(\sigma_{\delta\delta})$ if $p^* > 0$ and $p^* + dp^*$ is larger than $p^*(\sigma_{\delta\delta})$ if $p^* < 0$. Again, we see that a policymaker who seeks local robustness with respect to $\sigma_{\delta\delta}$ will follow a precautionary strategy by being less aggressive. More generally, if a policymaker's preferences are described by (24), then one can show from (25) that the introduction of a desire for robustness implies that 1) p^* is increasing in $E(\delta)$, 2) $|p^*|$ is decreasing in $\sigma_{\delta\delta}$, 3) p^* is decreasing in $\sigma_{\delta\upsilon}$ and 4) p^* is decreasing in r if $E(\delta) > 0$ whereas p^* is increasing in r if $E(\delta) < 0$.

Relative to these results, in particular eqs. (25) and (26), the introduction of multiple core models requires the replacement of model specific versions of $E(\delta)$, $\sigma_{\delta\delta}$, and $\sigma_{\upsilon_i\delta}$ by their counterparts as calculated via model averaging, as described by eqs. (6) and (7). Once one replaces the model-dependent moments into (25) with the moments described by (6) and (7), one can then proceed with various forms of robustness analysis.

Following our earlier discussion, we first focus on the variance of the policy variable coefficient. Let $\sigma_{\delta\delta, m}$ denote the variance of the policy coefficient conditional

on model m; the corresponding variance of the policy coefficient when one uses formula (7) to eliminate model dependence is $\sigma_{\delta\delta}$. Suppose that an adversarial agent chooses $\sigma_{\delta\delta,1}$ from the neighborhood $[\bar{\sigma}_{\delta\delta,1} - \Delta, \bar{\sigma}_{\delta\delta,1} + \Delta]$. Letting $\mu(m = i)$ denote the posterior probability of model i, one can show that

$$\frac{dJ}{d\sigma_{\delta\delta,1}} = -\frac{r}{2}\mu(m = 1)p^{*2} \tag{27}$$

This means that the least favorable variance for the policymaker is $\bar{\sigma}_{\delta\delta,1} + \Delta$. In response, the policymaker will adjust the policy variable according to

$$dp^* = \frac{-\mu(m = 1)p^*}{\bar{\sigma}_{\delta\delta}}\Delta \tag{28}$$

This equation is quite intuitive. It says that the policymaker will reduce the level of the policy variable and that this reduction is increasing in the degree of risk aversion, r, and in the probability of model 1.

One can also discuss robustness with respect to the model probabilities. For simplicity, we assume there are only two models. This allows one to assess robustness with respect to $\mu(m = 1)$ without having to specify where the change in probability for this model affects others (in the case of two models, changing the probability of one of the models of course means the other changes by an opposite amount.). Letting J_1 denote the policymaker's loss under model 1 and J_2 the loss under model 2, then

$$\frac{dJ}{d\mu(m = 1)} = J_1 - J_2 \tag{29}$$

This formula indicates if one is considering robustness with respect to posterior model probabilities in the interval $\left[\mu(m = 1) - \Delta, \mu(m = 1) + \Delta\right]$, the value against which one guards will depend on the relative values of J_1 and J_2. Suppose that $J_1 > J_2$, so that

30

model 1 is preferred by the policymaker, conditional on p^*. In this case, the optimal policy response will follow

$$\mathrm{d}p^* = \frac{1}{r\sigma_{\delta\delta}}[E_1(\delta) - E_2(\delta) - rp^*(\sigma_{\delta\delta,1} - \sigma_{\delta\delta,2}) - r(\sigma_{\delta\upsilon_i,1} - \sigma_{\delta\upsilon_i,2})]\Delta \tag{30}$$

where E_m denotes the expectation under model m.

In general, it is unclear whether the change described by eq. (30) is positive or negative. However, if r is small, then $\mathrm{d}p^* \approx -\left(E_1(\delta) - E_2(\delta)\right)\Delta$, so that if $E_1(\delta) > E_2(\delta)$, then the policy is used less aggressively when robustness is incorporated into the policy construction.

5. Issues in empirical implementation

In this section we turn from the theoretical side of model uncertainty to a discussion of how to incorporate model uncertainty into empirical exercises. This section discusses some operational issues; Section 6 will provide some empirical exercises.

i. Bayes, frequentist, or Wald approaches to model evaluation

From the perspective of empirical analysis, the key objects that must be computed are $\mu(\theta|d,m)$ and $\mu(m|d)$. These calculations require that a researcher take a stance on the use of Bayesian versus frequentist methods. In this section, we describe how this is so and show that the basic model averaging idea may be applied in both Bayesian and frequentist contexts.

a full Bayesian approach

The basic framework we have described corresponds to the way a Bayesian would model a decision problem, once one specifies a way of estimating $\mu(\theta|d,m)$ that formally accounts for prior information. To see this, notice that

$$\mu(\theta|d,m) = \frac{\mu(\theta,d|m)}{\mu(d|m)} = \frac{\mu(d|\theta,m)\mu(\theta|m)}{\mu(d|m)} \tag{31}$$

or

$$\mu(\theta|d,m) \propto \mu(d|\theta,m)\mu(\theta|m) \tag{32}$$

This latter formulation is the classic Bayes' rule. The key idea is that the description of uncertainty about θ given data d, also known as the posterior density, depends on two terms: $\mu(d|\theta,m)$, the probability of the data d given θ and $\mu(\theta|m)$, the probability of θ conditional on model m. Notice that in our interpretation, this prior density represents the uncertainty about θ that exists before the data d are realized. We do not assume that these unknowns are necessarily intrinsically random (such an assumption may not be appealing when the unknowns are parameters that characterize the economy, but is of course natural when the unknowns are shocks). Rather, the uncertainty about θ is subjective as it is characterized relative to the policymaker.

This formulation is what David Lindley[53] has called "The Complete Bayesian Paradigm," concluding

> Notice how constructive the paradigm is. It is like a recipe. You only have to follow the rules. What do you know?...What is uncertain?...What are the possible decisions?…In the coherent system, it is perfectly clear what has to be done. The difficulties are the evaluation of some of the probabilities and utilities and the calculation of others…

Lindley's distinction between evaluating and calculating probabilities alludes to a standard objection to the assumption in Bayesian methods that all uncertainty may be

described (evaluated) in terms of probabilities. This worry should not be dismissed; eminent statisticians such as David Freedman are not Bayesians for this reason. However, in our view, the correct response to this objection is to recognize that decisions on priors are perfectly defensible on pragmatic grounds. Eric Leeper, Christopher Sims and Tau Zha provide a good example of this and persuasively argue in favor of their use of informal methods to place prior restrictions on impulse response functions in order to produce plausible results. As these authors remark:[54]

> We could have accomplished the same, at much greater computational costs, by imposing our beliefs about the forms of impulse responses as precise mathematical descriptions, but this would not have been any more "disciplined."…There is nothing unscientific or unreasonable about this. It would be unscientific or dishonest to hide results for models that fit much better than the one presented…or for models that fit about as well as the one reported and support other interpretations of the data that some readers might regard as reasonable…

The basic message we wish to communicate is that accounting for model uncertainty can be done using standard Bayesian statistical methods.

model uncertainty and frequentist methods

While a full Bayesian approach provides a coherent way of dealing with model uncertainty, it does not constitute a unique strategy for doing so. The basic logic of treating the true model as an unknown and accounting for this can be readily adapted to frequentist data analyses; we will term this a pseudo-Bayesian approach. To see this, suppose that conditional on model m and data d, a policymaker assigns losses to each policy and data combination via some function $l(p|d,m)$. We interpret this as a frequentist loss function; the idea is that given a model and data, one may compute sample moments of interest to the policymaker and define losses with respect to them. This function may in turn be thought of as a random variable that has been conditioned on another random variable, namely model m. One can therefore eliminate this

dependence on m using the standard formula for conditional probabilities, i.e. compute an expected loss of the form

$$E\big(l\big(p|d\big)\big)=\sum_{m\in M}l\big(p|d,m\big)\mu\big(m|d\big)\propto\sum_{m\in M}l\big(p|d,m\big)\mu\big(d|m\big)\mu\big(m\big)\qquad(33)$$

While the last term of this expression requires a statement of prior probabilities on the model space, it does not require assigning prior probabilities to the unobservables contained in θ. From the perspective of frequentist calculations, $\mu\big(d|m\big)$ may be approximated by the standard likelihood statistic.[55]

   While an orthodox Bayesian might object to analyses such as (33) using the standard critiques of frequentist statistical methods, this is not relevant for our objective of providing ways to enhance the utility of empirical analyses of policies.[56] In our empirical applications, we shall use both full Bayesian and pseudo-Bayesian strategies to illustrate how each may be made operational.

Waldean approach

   Perhaps the major non-Bayesian approach to decision theory is due to Abraham Wald. In this type of analysis, the focus is on the development of statistical decision functions, i.e. the modeling of $p(d)$ which is a mapping from the space of data to the space of possible policy choices. The expected loss for a decision rule depends on the unknown θ. This leads to the notion of the risk function associated with a given statistical decision function:[57]

$$R(p,\theta)=\int_{D}l(p(d),\theta)\mu(d|\theta)\mathrm{d}d\qquad(34)$$

Policy rules are thus evaluated with respect to their associated risk. Risk functions, however, can only be evaluated conditional on θ. There are a range of ways to eliminate this conditioning when θ is unknown. If uncertainty about θ is described by a probability density $\mu(\theta)$, one can choose $p(d)$ so as to minimize expected risk

$$E\left(R(p,\theta)\right) = \int_{\Theta} \left(\int_{D} l(p(d),\theta)\mu(d\,|\theta)\mathrm{d}d \right) \mu(\theta)\mathrm{d}\theta\,. \tag{35}$$

By a standard calculation,[58] the evaluation of average risk leads to the same expected loss calculation as (1) when one uses the complete Bayesian solution we have described; $\mu(\theta)$ functions as a prior density.

A meaningful contrast between Wald and Bayesian approaches occurs if instead one follows a minimax strategy, i.e. choose $p(d)$ so as to minimize

$$\max_{\theta \in \Theta} \int_{D} l(p(d),\theta)\mu(d\,|\theta)\mathrm{d}d \tag{36}$$

Are there cases where the Wald approach can yield useful insights? The answer reduces, in our view, to the question of how one wants to handle priors and so must be handled in context. For example, in the Hansen-Sargent context where model uncertainty is defined around a single core model, the minimax strategy seems quite appealing. Similarly, our discussion of ambiguity aversion provides a justification for applying the Wald approach with respect to cross-model uncertainty regardless of how one evaluates within-model uncertainty.

ii. Characterizing model uncertainty

specifying elements of the model space

The specification of a space of possible models is ultimately a matter of a researcher's judgment. In one trivial sense, this follows whenever two researchers disagree on what models should be assigned zero prior probability. At the same time, our general view of disagreements in economics about models suggests that it is useful in specifying a model space to consider several distinct levels of model uncertainty and build up the space sequentially. The following levels are, we believe, a useful way to structure the building up of a model space.

– **Theory uncertainty.** As a rule of thumb, we would argue that model uncertainty occurs first because of theory uncertainty. Continuing disagreements among macroeconomists over the degree of price flexibility, the role of rational expectations and forward looking behavior in describing individual decisions, etc. are a good illustration of the limits to which the current state of economic theory can guide a policymaker; and of course, the persistence of disagreements over fundamental aspects of the economy reflects the absence of empirical evidence that is decisive in adjudicating alternative theories. At the same time, there are in most policy-relevant cases a rich range of alternative theories whose empirical analogs can form the first dimension along which to characterize the model space.

– **Specification uncertainty**. Once one has specified a range of theories, model uncertainty may then be discussed from the perspective of specification uncertainty. Standard examples of specification uncertainty in macroeconomic contexts include lag length for vector autoregressions and possible nonlinearities in the processes under study. Another form of specification uncertainty relates to measurement. In contexts such as growth economics, there are many empirical proxies that have been proposed for a given theory.

– **Heterogeneity uncertainty.** A third level of uncertainty in model specification concerns the extent to which different observations are assumed to obey a common model. In business cycle contexts, one needs to determine whether a model is rich enough so that data generated during a boom and during a recession may be interpreted as realizations from the same model. In growth contexts, one needs to determine the extent to which one allows for exceptionalism in the experiences of individual countries or regions. Different specifications of heterogeneity in turn produce different models.

To be clear, these levels of uncertainty are not "natural kinds." One can interpret heterogeneity uncertainty in many cases as a question of incorporating nonlinearity and

so can be interpreted as a form of specification uncertainty. Our purpose is exclusively to indicate some of the judgments that need to be made in constructing a model space.

interpreting a model space

While the specification of a model space is something that may only be discussed in the context of a particular economic phenomenon, a distinct issue is whether the analysis assumes that the "true" model is an element of the space. Jose Bernardo and Adrian Smith[59] distinguish environments that are M-closed and M-open; M-closed environments are those where the true model is unknown, but is included in the model space; in M-open environments, none of the models under analysis is true. From the perspective of model averaging procedures, as the number of observations increase, the "true" model will receive an asymptotic weight of 1 (so long as appropriate prior coefficient densities are used; see Fernandez, Ley and Steel[60] for discussion); when no model is true, that model which best approximates the data (in a particular sense based on Kullback-Leibler distance) will asymptotically receive a weight of 1.

While the asymptotics of statistical procedures that account for model uncertainty are reasonably well understood for both the M-closed and M-open cases, there has been relatively little work on the analysis of decision rules in M-open contexts. Bernardo and Smith propose some ways of engaging in statistical decision theory when no model is true; they do this in a very special context where the action of the modeler is the choice of a model and the objective of the modeler is the prediction of a future observation. The analysis unfortunately does not readily generalize to the sorts of problems which typically face economic policymakers, one reason being the question of interpreting counterfactuals in light of the Lucas critique; nor does the analysis address the model averaging approach we advocate.

The evaluation of policies in M-open cases is, in our judgment, an important open question. At the same time, we would note that the concern should not be overstated, at least in our context. Incorporating model uncertainty into policy analysis is the most appropriate way, we believe, in minimizing the role of misspecification in distorting policy evaluation. The objective of our model averaging approach is explicitly to treat

alternative models of the economy as potential candidates for the true model and allow the data to distinguish between them. Concerns about the absence of a true model in the space of potential models can thus only apply to models which the researcher has failed to foresee as a possibility. (The analysis of decision rules in the presence of unforeseen types of misspecification lies at the frontiers of decision theory as it requires thinking about decisions when the decisionmaker does not know the support of the uncertainty he faces. While some aspects of this problem have been addressed in recent work in economic theory, it is far from well understood.) Further, since the specification of a model space will presumably evolve over time as more information becomes available to an analyst, at least asymptotically the assumption that the space is M-closed may not be as strong as it first appears.

specifying prior probabilities on models

A final issue in characterizing model uncertainty concerns the construction of prior probabilities over models. The specification of prior probabilities on a model space raises many conceptual issues. Some of these issues are related to the general questions concerning the nature of prior probabilities that continue to be debated in Bayesian contexts. Our own views in this regard are pragmatic. Desiderata in the assignment of priors include, in our view

– **informativeness with respect to the likelihood.** Priors should assign relatively high probability to those areas of the likelihood that are relatively large. Otherwise, the prior will have excessive impact on the posterior description of the parameters.[61]

– **robustness**. A prior should be robust in the sense that a small change in the prior should not induce a large change in the posterior. As argued by James Berger,[62] robustness may be interpreted as a safeguard against misspecification of prior information.

– **ability to serve as benchmark.** Priors should be flexible enough to allow for their use across similar studies and thereby facilitate comparability of results.

Of course, these obviously desirable properties leave a great deal of discretion to a given researcher. And one can easily add other desiderata to our list. The arguments made by Leeper, Sims, and Zha,[63] described above, suggest that "reasonableness of results" should be included. This lack of algorithmic precision in the assignment of priors is in our view appropriate; priors ultimately are at least partially a nuisance whose choice should be regarded as nothing more than facilitating the presentation of salient features of the data.

How do these simple principles apply to the model uncertainty context? At first glance, it might seem that if one does not have such information, one should assign equal prior weight to each element of M. However, this is not entirely satisfactory as it ignores interrelations between different models.

The problem is easiest to see in the case of linear regression models. Suppose that one is considering model uncertainty where different models correspond to different choices of which control variables to include in a linear regression. This is the problem described in the context of eq. (9) and one to which we will return in the context of growth econometrics in Section 6. The recent efforts to employ model averaging to account for uncertainty with respect to variable inclusion[64] generally assume that the possible models are all equally likely a priori. So, in the case of linear regressions where there is uncertainty over which of K regressors are present, each of the 2^K models in the model space is assigned probability 2^{-K}. This is equivalent to assuming that the probability that a given variable is present in the "true" model is equal to .5 and is independent of the presence or absence of any of the other regressors in the model. Proposals have been made to alter the probability of variable inclusion in order to give greater weight to models with a small number of regressors,[65] as well as to assume the probability that a given variable is included is itself a random variable drawn from some distribution, thereby allowing different variables to be included with different probabilities,[66] but the independence assumption is, at least in our reading, essentially universal.

As argued by Brock and Durlauf,[67] such a formulation of priors on the model space is difficult to justify. The growth theory that the rule of law affects growth may be logically distinct from the theory that property rights affect growth, but that does not mean that the fact one matters has no implications for the likelihood that the other does. This problem is thus closely related to the red bus/ blue bus paradox that appears in discrete choice theory. The discrete choice question is how the probability an individual chooses a red bus or a taxi is affected by the addition of the possibility of using a blue bus as well. Under the independence of irrelevant alternatives assumptions of a logit model, the presence of the blue bus should not affect the ratio of the choice probabilities between a red bus and a taxi; this is an unappealing feature since the blue bus is a far closer substitute to the red bus than the taxi. The discrete choice literature has proposed a number of ways of addressing these types of issues, including nested logit models, which organize choices in a tree structure that reflects similarities (modeled in the nested logit context as common utility components). We will use an analogous approach in defining model probabilities in the applications we take up next.

6. Empirical applications

i. Monetary policy rules

Our first example concerns monetary policy rules and is designed to illustrate a way of integrating model uncertainty using frequentist (or what we called pseudo-Bayesian) methods. The last decade has seen an explosion of research on alternative policy rules, much of it stimulated by the seminal work by John Taylor on what is now called the Taylor rules. In this section, we present some results on Taylor rules and model averaging. For simplicity, we use a conventional loss function that is quadratic in output, inflation and interest rates; assume monetary policy is constrained to follow a Taylor rule; further, we only consider backward looking models. We compute robust estimates of the effects of alternative choices of monetary policy parameters. We contrast those estimates with those of the well-known Rudebusch and Svensson model.[68]

Model uncertainty has played a prominent role in recent analyses of monetary policy. An early example is Bennett McCallum's analysis of normal income rules, which experimented with alternative Phillips curve specifications in order to establish robustness across results.[69] The same concern with robustness appears in a number of papers in the Taylor volume, and in recent papers such as Levin and Williams.[70] Like much empirical research, this literature typically proceeds on the intuition that the set of estimates produced will bracket the actual effect of a policy under consideration (or, more modestly, is likelier to bracket the effect than is a set produced by extrapolating results from a single model).

As explained above, what we offer is a procedure for formally combining the estimates from a set of models. In this section, estimates are weighted by the corresponding model's likelihood (adjusted for degrees of freedom) and by prior model probabilities. We set these prior probabilities equal for all models, so weights are simply the model likelihood: well-fitting models get more weight than do ill-fitting models. We view our approach as a complement rather than replacement for that described in the previous paragraph. Formal model combination will help focus attention on a central tendency across models. But economists and policymakers will still find it useful to answer the question, "if one puts prior weight of unity on one or another model, what is the risk?"

The approach that we have proposed is well suited to consider what may be the central source of such uncertainty in monetary policy analysis, namely, the modeling of expectations. We share the view of many economists that explicit modeling of expectations is relatively important when one is considering the effects of a permanent change in regime, say a switch to inflation targeting. Models with an atheoretical lag structure are relatively appealing if one wants to think about the tradeoff between (say) raising interest rates 50 basis points this month, or 25 basis points this month and 25 basis points next month, when either action is within the framework of how monetary policy is currently conducted. Our approach naturally accommodates this view, by allowing one to choose model weights (choose $\mu(m)$'s) that vary with the question at hand.

In this first analysis, however, we limit ourselves to models in which expectations are backward looking. Indeed, we abstract from simultaneity of any sort even that

associated with Cowles Commission style simultaneous equations models. With various definitions of "robust," but also with the use of quadratic preferences, Taylor rules and backward looking models, calculations similar to ours have been supplied by Alexei Onatski and James Stock and Onatski and Noah Williams.[71] The research presented here is intended to both complement this work and to illustrate the frequentist approach to model averaging (eq. (33)) in a simple context.

We employ the same notation as Section 4: y_t is the output gap; π_t is the quarterly inflation, at annual rates; i_t is the federal funds rate, $\bar{\pi}_t = \frac{1}{4}\sum_{j=0}^{3}\pi_{t-j}$, and $\bar{i}_t = \frac{1}{4}\sum_{j=0}^{3}i_{t-j}$. We assume that policymakers wish to minimize

$$R = var\left(\pi_t\right) + \lambda_y var\left(y_t\right) + \lambda_i var\left(\Delta i_t\right) \tag{37}$$

Following the literature, R is referred to as a measure of the risk of a policy. We do not attempt to link parameters to a particular microeconomic model,[72] nor do we allow the weights to vary across specifications.

We consider three equation models for i_t, y_t, and π_t Our specification assumes that the output gap and inflation rate are predetermined. The nominal interest rate is determined by a Taylor rule

$$i_t = g_\pi \pi_t + g_y y_t + g_i i_{t-1} \tag{38}$$

In (38) and elsewhere we suppress constants and all other deterministic terms.

We consider models in which the output gap y_t and quarterly inflation π_t depend on up to four lags of i, y, and π. We label the equation with y_t on the left hand side as the IS curve and the equation with π_t on the left hand side as the Phillips curve. The right hand side of the IS equation always includes at least one lag of y and one lag of an annual or quarterly ex-post real interest rate, although we do not in all specifications constrain coefficients on nominal interest rates and inflation to be equal and opposite.

The right hand side of the Phillips curve equation always includes at least one lag of inflation and one lag of output, with the lags of inflation constrained to sum to unity. The most profligate specification entailed four lags of i, y, and π in both equations, which was almost but not quite an unrestricted VAR ("almost" because lags of inflation in the Phillips curve were always constrained to sum to 1).

Specifically, we varied lags across specifications as follows. In the IS curve, we included specifications of two types. First, we constructed specifications with a single lag of the annual ex-post real interest rate $\bar{i}_{t-1} - \bar{\pi}_{t-1}$ along with alternative lags for y of : lag 1, lags 1-2, lags 1-3, and lags 1-4; lags for π of: none, lag 1, lags 1-2, lags 1-3 and lags 1-4; and lags for i of: none, lag 1, lags 1-2, and lags 1-3. This set of $4 \times 5 \times 4$ alternative specifications may be written as

$$y_t = \alpha_{y1} y_{t-1} + \alpha_{r1}\left(\bar{i}_{t-1} - \bar{\pi}_{t-1}\right) + \left[\sum_{j=2}^{4} \alpha_{yj} y_{t-j} + \sum_{j=1}^{4} \alpha_{\pi j} \pi_{t-j} + \sum_{j=1}^{3} \alpha_{ij} i_{t-j}\right] + u_t \qquad (39)$$

The first two terms on the right hand side of (39) were included in all regressions. The terms in the brackets describe the additional regressors. Additional IS specifications were obtained with models that are identical to those we have just described, except that a single lag of the quarterly ex-post real interest rate, $i_{t-1} - \pi_{t-1}$, was always present, with lags of i adjusted to prevent linear dependence in the regressors in particular versions of (39). This also produces $4 \times 5 \times 4$ specifications.

In the Phillips curve, specifications included lags for y of: lag 1, lags 1-2, lags 1-3, and lags 1-4; lags for π of: lag 1, lags 1-2, lags 1-3, and lags 1-4; and lags for i of none, lag 1, lags 1-2, lags 1-3 and lags 1-4. This set of $4 \times 4 \times 5$ specifications may be written

$$\pi_t = \beta_{\pi 1} \pi_{t-1} + \beta_{y1} y_{t-1} + \left[\sum_{j=2}^{4} \beta_{\pi j} \pi_{t-j} + \sum_{j=2}^{4} \beta_{yj} y_{t-j} + \sum_{j=1}^{4} \beta_{ij} i_{t-j}\right] + v_t, \quad \sum_{j=1}^{4} \beta_{\pi j} = 1 \qquad (40)$$

Once again, the first two terms on the right hand side of (40) were included in all regressions and the terms in brackets describe the additional regressors.

Each of the regressions we have described was estimated alternately with a constant term as the only deterministic component and with a constant term as well as a

43

post-1984:I dummy. The dummy is intended to crudely allow for changes initially documented by Margaret McConnell and Gabriel Perez-Quiros[73]. Thus the total number of specifications is $\left(\left(4\times5\times4\right)+\left(4\times5\times4\right)\right)\times4\times4\times5\times2=25,600$ where the final "2" accounts for the two sets of deterministic terms.

In all computations, we discarded specifications whose estimates implied behavior that was nonstationary. Mechanical processing of such estimates would yield unbounded variances and infinite risk. Our view is that in a full treatment such estimates should be dampened to yield finite variance and risk, in accordance with our prior knowledge that the output gap and inflation are stationary. Discarding the estimates was done for simplicity.[74]

For each model, we estimate the IS and Phillips curves by least squares. In conjunction with choices of g_π, g_y, and g_i in (38), one can compute estimates of the total loss described by (37) using point estimates of the variances implied by the model. For model m, we refer to this estimated loss as $\hat{R}_m$. For each model, we compute a BIC-adjusted likelihood, call it L_m. We compute model average risk as

$$ER = \frac{\sum_{m\in M} \hat{R}_m \mathrm{L}_m}{\sum_{m\in M} \mathrm{L}_m} \tag{41}$$

This equation fits into the frequentist approach outlined in section 5 with $\hat{R}_m$ playing the role of $l\left(p|d,m\right)$ and $\dfrac{\mathrm{L}_m}{\sum_{m\in M}\mathrm{L}_m}$ the role of $\mu\left(d|m\right)$ in eq. (33), under the assumption that all models have equal prior probabilities, i.e. $\mu\left(m\right)=\dfrac{1}{25,600}$.

To clarify and illustrate the effects of model averaging, we contrast the model averaging results with those of one well-known special case of the class of models considered. This is the Rudebusch and Svensson model.[75] In this model, the IS equation is

$$y_t = \alpha_{y1}y_{t-1} + \alpha_{y2}y_{t-2} + \alpha_{r1}(\bar{i}_{t-1} - \bar{\pi}_{t-1}) + u_t \tag{42}$$

and the Phillips Curve equation is

$$\pi_t = \beta_{\pi1}\pi_{t-1} + \beta_{\pi2}\pi_{t-2} + \beta_{\pi3}\pi_{t-3} + \beta_{\pi4}\pi_{t-4} + \beta_{y1}y_{t-1} + v_t \tag{43}$$

where $\sum_{j=1}^{4}\beta_{\pi j} = 1$ is imposed so that the long run Phillips Curve is vertical.

For a range of values of parameters λ_y and λ_i in the risk function (37), we solved for Taylor rule parameters that were optimal under Rudebusch and Svensson. We computed risk according to the Rudebusch and Svensson model, denoted as $\hat{R}_{RS}$, as well as according to all other models in the model space we have described. The model specific risk calculations were then averaged according to (12) to produce average model risk. The objective of this exercise is to see whether the Rudebusch and Svensson figures for risk well match those for model averages. The range of values for the risk parameters were those suggested by Levin and Williams,[76] $\lambda_y = \{0.0, 0.5, 1.0, 2.0\}$ and $\lambda_i = \{0.1, 0.5, 1.0\}$, 12 sets of values in all.

Apart from lags, the sample is 1969:I-2002:IV. Inflation is computed as annualized growth in the GDP deflator, the output gap is computed from real GDP and the Congressional Budget Office (CBO) estimate of potential GDP. We used the latest data available, thus abstracting from possible complications from data revision.

Results are in Table 1. Columns (g_π), (g_y) and (g_i) give values of the Taylor rule parameters that are optimal under Rudebusch and Svensson, found by a grid search. These display a familiar and intuitive pattern. Higher weights on output volatility (higher λ_y) lead to higher optimal g_y, higher weights on interest rate volatility (higher λ_i) lead to higher optimal g_i. As has been found in previous studies of the Rudebusch and Svensson model, the optimal interest rate parameter g_i is not very large, and sometimes is negative.

For the Taylor rule parameters given in the Table, we compute model average risk ER based upon eq. (41) and compare it to the Rudebusch-Svensson risk $\hat{R}_{RS}$. In principle, model average risk can be higher or lower. And indeed we see that the last column of Table 1 includes both negative and positive values, with positive values indicating that model average risk is higher. Relative to Rudebusch-Svensson risk, model average risk tends to be high where there is a relatively small penalty to interest rate volatility and low when there is a large interest rate penalty. While the last column figure in the first line of the Table is quite large, the other numbers are much smaller and scattered fairly evenly around zero.

We take this as illustration of two points. First, upon comparing our results with those of Levin and Williams,[77] it seems that there is substantially less variation in risk within the class of backwards models we have studied than there is between backward and forward looking models. Specifically, findings for the Rudebusch-Svensson baseline is generally representative of the risk associated with the monetary policies considered in the table. Second, and potentially more useful from the perspective of future research, is one emphasized in our discussion above: model averaging allows tractable accounting for the effects of model uncertainty.

ii. economic growth

In our second application, which will follow the full Bayesian approach we discussed in Section 5, we turn to the empirical growth literature. Our analysis will focus on the evaluation of the effect of tariffs on economic growth. In order to develop the empirical exercise, we first discuss some general issues in growth econometrics.[78]

growth econometrics: general issues

Much of recent macroeconomic analysis has focused on issues associated with economic growth. The empirical basis for much modern growth research is the now classic cross-country growth regression.[79] From the vantage point of using such regressions to evaluate a growth policy p, a canonical form of this regression is

$$g_i = \beta' X_i + \gamma' Z_i + \delta p_i + \varepsilon_i \qquad (44)$$

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 (initial 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 importance of such regressions in policy analyses is summarized by Edmond Malinvaud[80]

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

Regressions such as (44) have been used to evaluate many policies: a survey of this type of empirical work may be found in Barro and Sala-i-Martin.[81] For our purposes, the main point is that the evaluation of a growth policy typically amounts to assessing the statistical significance of δ for a baseline specification (44) and a small set of alternative specifications which typically amount to changing the variables that are included in Z_i. Such analyses pay only indirect attention to the question of the space of models and how to evaluate differences across models in drawing conclusions about parameters of interest.

From the perspective of evaluating growth policies, this standard approach may be faulted using arguments we have developed.[82] One problem is that the choice of control variables to include as components of Z_i is typically very ad hoc. A survey by Durlauf and Danny Quah[83] found nearly as many alternative growth theories and associated empirical measures as there are countries in the standard data sets; by now the number of theories exceeds the number of countries. This plethora of alternative theories

is particularly worrisome because, following Brock and Durlauf,[84] growth economics suffers from theory openendedness. Theory openendedness means that one growth theory typically has no logical connection to the empirical possibility of another. The theory that political stability affects growth is compatible with any number of other theories, such as the claim that the composition of natural resources affects growth.

Second, empirical growth research has generally not dealt systematically with questions of heterogeneity in the growth processes for different countries. Regressions such as (44) are interpretable for policy evaluation only to the extent that one believes that the regression specification is sufficiently rich that the data from each country constitutes a draw from the common statistical model defined by the regression. While this requirement is hardly unique to growth contexts, its plausibility is particularly questionable when one is working with such complicated objects as national economies. To be concrete, suppose that one wishes to advise a country on some policy using a cross-country regression as a source of empirical evidence. Does one believe that the growth implications of a unit change in a given policy variable is the same for the United States as countries in sub-Saharan Africa? It is easy to think of cases, for example changes in the percentage of high school graduates in the labor force, where one would not wish to make such an assumption, but this is precisely what is asserted when one uses (44) to uncover growth determinants.[85]

There are a number of studies[86] that have documented parameter uncertainty of various forms. The sorts of parameter heterogeneity that have been identified have often been interpreted to indicate how different stages of socioeconomic development are associated with different growth processes. Even if one does not believe that the empirical case for parameter heterogeneity has been established, there is certainly enough such evidence to allow for the possibility in policy evaluation exercises.[87]

A third problem is that it is far from clear that statistical significance can provide a useful guide to policy evaluation. While the abstract argument was made in Section 2, it is particularly salient in the growth context and so we expand upon it. Suppose the purpose in using linear growth regressions is to evaluate whether country i should make the policy change from $\bar{p}$ to $\bar{\bar{p}}$. As we have suggested earlier, standard practice in the growth literature is based on the use of the t-statistic associated with $\hat{\delta}$ to evaluate the

policy. Following Brock and Durlauf,[88] one can think about t-statistics rules from a decision-theory perspective. A simple way to do this is to interpret a t-statistic rule as implying that, when comparing $\bar{p}$ versus an alternative policy $\bar{\bar{p}} > \bar{p}$, one will only move from $\bar{p}$ to $\bar{\bar{p}}$ if the associated t-statistic for the policy parameter δ is greater than 2. Further, interpreting a t-statistic as the ratio of the mean of the parameter to its standard deviation, one can approximate the t-statistic rule as implying that one makes the policy change based upon

$$E\left(l\left(\bar{\bar{p}},\theta\right)\middle|d,m\right) - E\left(l\left(\bar{p},\theta\right)\middle|d,m\right) = \left(-E\left(\delta\middle|d,m\right) + 2\operatorname{var}\left(\delta\middle|d,m\right)^{1/2}\right)\left(\bar{\bar{p}} - \bar{p}\right) \quad (45)$$

with the policy change taken only if the value of (45) is less than 0. (If we are considering a reduction in the policy variable, the requirement would be that (45) is greater than 0; this will be relevant when we consider the question of a tariff reduction.) This is a special preference structure in two senses. First, it assumes that one's evaluation of the policy depends on the effect of the policy on growth and not on growth itself. Second, it assumes a very particular tradeoff between the mean and variance of the policy effect.[89]

This interpretation of the t-statistic rule may also be used when one has averaged across models; one simply computes the formula using moments on the right hand side of (45) that are conditioned on the data d but not on a specific model m. We will use this below to facilitate comparisons between policy advice for different models and model averaging.

evaluating a policy of tariff reduction to enhance growth

In order to show how one might address these problems we consider a particular policy question: should the countries of sub-Saharan Africa[90] lower tariffs in order to improve growth performance? Our analysis is obviously a caricature of the actual policy process as it ignores the plethora of information that is available to organizations such as the World Bank that help inform policy decisions, but for expositional purposes we treat cross-country growth regressions as the sole basis on which policy decisions are made.

In order to evaluate this policy question, we proceed as follows. First, we define a set of different growth theories that have been proposed in the empirical literature. This constitutes a first level of model uncertainty. Second, for each theory, there is uncertainty as to which empirical proxies to employ to capture it. Third, we allow for uncertainty concerning whether sub-Saharan Africa countries obey the same growth process as the rest of the world.

With respect to theory uncertainty, we proceed as follows. In every model, we include the variables predicted by the Solow growth model and our tariff variable. With respect to eq. (44), this means that every element in the model space contains X_i and p_i.[91] We then introduce six possible additional categories of theories of cross-country growth differences that have received prominence in the literature: 1) exchange rate policies, 2) government spending policies, 3) inflation, 4) characteristics of the economic system, 5) characteristics of the financial system, and 6) characteristics of the political system. The first three categories, roughly speaking, represent theories that relate various government policies to economic growth. The second three categories represent theories that link growth to longer run structural aspects of a country's economic and political system. While these categories do not exhaust the range of new growth theorizing, we do argue that they contain a relatively comprehensive range of new growth theories.

The construction of this first stage of the model space for cross-country growth behavior requires a number of decisions on the part of the analyst. One decision concerns the ways in which alternative theoretical specifications are defined. We interpret each theoretical specification for a growth model as the choice of a set of theories to include in a growth regression. We therefore rule out combinations of theories as would occur if one were to use the space of empirical growth proxies to recombine elements as is done in factor analysis. Such alternative approaches are not, in our view, interpretable as growth models. However, there may be an argument for doing so in policy evaluation contexts, if one is indeed interested only in posterior distributions of policy effects; we defer this consideration to future work. Further, even if one restricts oneself to distinct theories, there are questions of how to organize variables into distinct theoretical categories. Our choices for distinct growth theories have been made in a way that we believe minimizes the connections across theories in the sense that one can treat the

probabilities of each theory being included as approximately independent. This is admittedly a judgment call, but is no different from the judgments often necessary to implement models such as the nested logit.[92]

Second, once one has specified a set of theories, it is necessary to specify how the various theories are characterized empirically. For each theory, we have identified a small number of variables that have been employed in the empirical growth literature to capture the theory; these various data series are defined in the Data Appendix. For each of these sets of variables, we allow each non-empty subset to correspond to a way of empirically modeling the theory. For example, for the theory that political structure affects growth, we have two empirical proxies: civil liberties, and an index of democracy. There are three different nonempty subsets of these variables that may be used to empirically instantiate the theory. Each subset choice corresponds to a distinct growth model.

Third, we model parameter heterogeneity in a way that allows us to treat it as a variable inclusion problem. Specifically, we use a very standard procedure in empirical work in that models with parameter heterogeneity will take the form

$$g_i = \beta' X_i + \gamma' Z_i + \delta p_i + \bar{\beta}' X_i \xi_{i,SSA} + \bar{\gamma}' Z_i \xi_{i,SSA} + \bar{\delta} p_i \xi_{i,SSA} + \varepsilon_i, \qquad (46)$$

where $\xi_{i,SSA}$ is an indicator variable that equals 1 if country i is in sub-Saharan Africa and 0 otherwise. This type of heterogeneity has proven useful in previous work on sub-Saharan Africa, cf. Brock and Durlauf, which found, reexamining an important study by William Easterly and Ross Levine that the effects of ethnic heterogeneity on growth are much stronger for Africa than for the rest of the world.[93]

Figure 1 illustrates our formulation of model uncertainty for growth regressions. The first level of uncertainty that must be resolved in defining a particular model concerns the set of growth theories to include in the specification. The second level of uncertainty that must be resolved is which empirical proxies for these theories are used. Once a set of theories and associated empirical proxies are specified, the final level of uncertainty that must be resolved is whether sub-Saharan Africa obeys a different growth process from the rest of the world or not. If one were to enumerate every sub-branch for

Figure 1, the final nodes would denote the universe of possible models. The levels of the tree indicate the levels at which we assign model probabilities; at each level probabilities are assigned equally to all possible branches. This procedure partially addresses the red bus/blue bus problem we described earlier.

This tree structure provides the basis on which we assign probabilities. With respect to theory inclusion, we assume that the inclusion probabilities are equal and that the theory inclusion probabilities are unaffected by what additional theories are included. This means, for example, that the probability that the "exchange rate" theory of growth appears in a model is independent of whether the "political structure" theory of growth is included in that model. Second, we assigned equal probability weights to each of the possible empirical analogs of a theory (i.e. to each combination of variables used to measure the theory). Third, for each specification of theories and associated variables, we specify versions with and without sub-Saharan African heterogeneity. Models with heterogeneity correspond to eq. (46); we allow the error variances for SSA countries to differ from the rest of the world. For each pair of corresponding models with and without heterogeneity, we assign probabilities q to the heterogeneous model and $1-q$ to the homogeneous model. For expositional purposes, we report $q = 0$ separately. Overall, there are 4096 different models generated by theory and regressor choice uncertainty; allowing for heterogeneity uncertainty doubles this to 8192.

This tree structure for the probabilities represents an effort to address the problem in previous work[94] that two empirical proxies for the same theoretical property are treated in the same way as two proxies for different theories in terms of their joint probabilities of inclusion. Our approach is designed to distinguish the questions of uncertainty over theories from questions of uncertainty concerning empirical proxies. While our approach is, we believe, an improvement on previous ways of assigning prior probabilities, we fully expect that it will evolve in future work.[95]

In order to compute posterior densities, for the parameters and associated expected growth levels in the models defined by (44) and (46), it is necessary to specify prior distributions on the model coefficients and a distribution on model errors. We assume a uniform prior on the coefficients and a Gaussian error distribution. As explained in the appendix, this has the important benefit that the posterior expected value

of the regression coefficients in a given model may be approximated by the OLS estimate of the parameters, and the posterior variance may be approximated by the OLS estimate of the parameters' variance covariance matrix. This makes our results straightforward to interpret from a frequentist perspective. However, we wish to be clear that this choice of priors is made primarily for expositional clarity; see Fernandez, Ley and Steel[96] for an extensive discussion of the appropriate use of priors in linear model averaging contexts.

Table 2 reports the results of our estimates of the posterior mean and standard deviation for the tariff parameter under a range of specifications. The tariff variable measures tariffs on intermediate goods and inputs and corresponds to OWTI in the standard Barro and Lee data set. Column (OLS) reports OLS estimates of the tariff variable based on regressions that include the Solow variables and the tariff variable. Column (Full OLS) reports OLS estimates when all possible variables are included. Columns (BMA) report Bayesian model averaging exercises under different theory inclusion probabilities q, we consider $q = .25, .5,$ and $.75$ respectively. Columns (Min Coefficient) and (Max Coefficient) report estimates for all models estimated in the BMA analysis that produce the minimum and maximum posterior means of the parameter. Columns (Min Mean $+2\sigma$) and (Max Mean $+2\sigma$) report the results for the analogous models whose payoffs under the t-statistic rule eq. (45) are minimal and maximal. The OLS regressions are included to serve as benchmarks in indicating where model averaging matters. (Recall that under our assumption, the OLS regression estimates of coefficient and associated standard errors correspond to the posterior means and standard deviations of the parameters, thus the OLS regression is a degenerate model averaging exercise, i.e. one where all prior model probability is assigned to one model. Columns (Min Coefficient) through (Max Mean $+ 2\sigma$) are useful in understanding how data mining and ambiguity aversion may be evaluated.

Table 2 indicates that estimates of the posterior densities of the parameters associated with the tariff variable are each very robust with respect to model uncertainty. The alternative probabilities of theory inclusion in Columns (BMA) have very little effect on posterior means and standard deviations. Relative to the OLS regressions in Columns (OLS) and (Full OLS), the model averaging estimates of the mean of the tariff parameter is more than 10% higher than the narrow Solow model and about 3% larger than the

expanded Solow model. For the standard deviation, our model averaging estimates are about 10% smaller than the narrow Solow model and similar to our estimate for the expanded Solow model. Notice that in each case the tariff variable is negative with a standard deviation less than half the size of the coefficient; by the "t-statistic" loss function described by eq. (45), these regressions would support the recommendation of a tariff reduction. Overall, the support for the policy change under these preferences is somewhat stronger when the posterior probabilities are computed using model averaging versus the OLS model. To be clear, the model averaging analysis does not lead to a different view of the policy advice suggested by the two OLS specifications. Its value added comes in showing that this advice is not an artifice of the choice of specification.

One can compare the model averaging results to those obtained under models that are singled out because they are particularly favorable or unfavorable for a policymaker with t-statistic preferences. If the policymaker is risk neutral, Column (Min Coefficient) reports the model that would provide the strongest support for a tariff reduction as it has the smallest parameter. A policymaker with t-statistic preferences would find the model described in Column (Min mean + 2σ) most favorable. We call these cases data mining models because an advocate of a tariff reduction would want to use these specifications in an effort to persuade the policymaker to implement the reduction. A policymaker who possessed an ambiguity aversion parameter $e = 1$ but only cared about the mean of the parameter conditional on a model would make a policy evaluation on the basis of the model described in Column (Max Coefficient) whereas an ambiguity averse policymaker with t-statistic preferences conditional on a model would evaluate a tariff reduction on the basis of the model described in Column (Max Mean + 2σ).

These results indicate that the policy recommendation that is implied by the OLS and model averaging exercises is similar to that which is implied by the data mining models. This occurs because models in the vicinity of the data mining models are associated with relatively large posterior probabilities. So, in this sense the support for the reduction is strengthened. In contrast, an extremely ambiguity averse agent will find the evidentiary support for the reduction to be far weaker. However, if the policymaker is risk neutral within a model, he will still conclude that the reduction is justified. The

policymaker with t-statistic preferences will not favor the reduction, but the payoff differential between the status quo and the reduction is not particularly large.

Table 3 extends our analysis to allow for heterogeneity between sub-Saharan Africa and the rest of the world. We report OLS estimates for (44) for the tariff parameter from regressions based on 1) the Solow variables plus tariffs and 2) the Solow variables, tariffs, and all other variables in Columns (OLS) and (Full OLS) respectively. For the model averaging analysis, we focus on the case where the theory inclusion probability is .5 and consider prior probability weights on models with heterogeneity and corresponding models without heterogeneity to equal .5, .75, and 1 respectively. These results appear in Column (BMA Prior Heterogeneity Probabilities).

Column (BMA Prior Heterogeneity Probabilities) indicates a significant discontinuity in the mean and standard deviation of the tariff parameter for $q = 1$ versus the other cases. In particular, the first two moments of the parameter are similar to those found in Table 2 for $q = .5$ and $q = .75$; allowing for heterogeneity slightly lowers the posterior mean and raises the posterior standard deviation by about 20% for the prior heterogeneity probability of .5 and by about 50% for the prior heterogeneniety probability of .75. In contrast, the posterior mean and standard deviation for $q = 1$ are very different; the mean is nearly doubled and the standard deviation is about 4 times as large as found for the model averaging counterparts in Table 2. The reason for the large differences is that the posterior probabilities on the subset of models that allow for SSA heterogeneity are very small. When $q = .5$, the total posterior probability on models with heterogeneity is only .014; for $q = .75$, the posterior probability is only .04. As a result, these models have relatively little effect on the overall posterior density of the tariff parameter. In contrast, $q = 1$ imposes heterogeneity on all models. This leads to the very different estimates that would, by preferences such as (45), lead a policymaker to advise against a tariff reduction. Our other regression exercises also lead to a rejection of the tariff reduction under preferences (45). In both of the Solow cases, if SSA heterogeneity is included with probability 1, the standard deviation of the posterior density of the tariff coefficient for SSA countries swamps the posterior mean. These results of course mean that a sufficiently ambiguity averse agent would not lower tariffs. A data miner could

produce a model, however, that supports a tariff reduction, as indicated by the most favorable models we report.

We are surprised by the weakness of the evidence on heterogeneity given previous work[97] that found parameter heterogeneity, albeit in a very different statistical context. However, the bottom line of this exercise is that sub-Saharan African heterogeneity does not appear to be important in the interpretation of our exercises with respect to policy evaluation except under a very high degree of ambiguity aversion.

As we have suggested in our earlier discussion of policy evaluation as a decision theory problem, using hypothesis tests to analyze growth policies suffers from the problem that statistical significance (or its analog) may not constitute an appropriate way to think about policymaker preferences. We therefore provide some additional analyses that allow one to discuss a tariff change as a counterfactual from the perspective of the distribution of growth rates. Table 4 reports an exercise for the sub-Saharan African economies in which the mean and variance of the growth rate for each country between 1960 and 1985 is compared with and without a 10% reduction of tariffs as compared to what occurred historically. To do this, we use the posterior means and variances of the model parameters β, γ, and δ based on the historical data. We then compute the posterior mean and variance of g_i with and without a 10% reduction in the tariff variable, keeping all other regressor values constant. We assume that the errors in the growth process are independent of the regressors. This type of exercise may be criticized using Lucas critique-type arguments, as we do not account for the effects of the policy change on model parameters (or for that matter on the other regressors which are themselves endogenous). Nevertheless, we think the exercise is useful in terms of illustrating how a decision-theoretic approach to evaluating the tariff policy differs from the conventional hypothesis testing approach. We also compare these estimates with those models which possessed the largest and smallest tariff coefficients. For the model averaging exercises, we employ a theory inclusion probability $q = .5$, which reflects our judgment that the theories we have allowed for are ex ante quite plausible, i.e. that the growth process is best understood as driven by a relatively large number of factors; we have separately verified that the results we report are quantitatively similar for other probability choices. We do not allow for parameter heterogeneity; as one would suspect from Table 3,

introducing such heterogeneity does not affect the findings if the prior heterogeneity property is .5 or .75. In addition to the model averaging exercises, Table 4 also reports results for the models with the largest and smallest tariff parameters.

What sorts of conclusions might one draw from the information in Table 4? One finding of importance is the heterogeneity in the expected growth levels across countries. Focusing on the estimates under model averaging, Botswana, for example, is associated with an expected growth over this period of over 100% whereas Burundi had an expected growth of -9%. The differences in the standard deviations are much smaller. The reason for this is that the uncertainty in the growth rates is very much dominated by the contribution of the model error. Even with these similarities in the standard deviations, the cross-country heterogeneity in the posterior densities of growth rates means that in general, one cannot make strong policy statements for mean/variance loss functions without explicitly calculating the moments of the growth process; the invariance of policy advice that one finds using a loss function such as (45) is not general. It is easy to construct loss functions where one would advise one sub-Saharan African country to lower tariffs but not another, using the same econometric information from cross-country growth regressions.

A second finding is that the effects of a change in tariffs on the standard deviation of a country's growth are far smaller than one would guess from looking at the standard deviation of the density for the tariff parameter in isolation. In fact, in many of the cases, one finds a reduction in the posterior standard deviation of the expected growth rate. The reason for this is that the different growth determinants may be interpreted as different elements of a portfolio; in the growth case they apparently act to reduce the overall variance of the growth rate, at least in terms of the data we have analyzed. This once again suggests the importance of specifying priors and computing posterior densities of the outcomes of interest, and not focusing on model parameters in isolation. From the perspective of a policymaker with mean/variance preferences, a tariff reduction may have desirable effects in terms of stabilizing the growth rate. These findings are not affected by considering the two extreme models reported in the Table.

In evaluating the results in Table 4, it is essential to keep in mind that the counterfactual assumed that the values of all the growth determinants X_i and Z_i are

known; so that all uncertainty about the growth process is generated via the parameters associated with the determinants. So, we certainly do not wish to argue that the estimates of variance in the expected component of growth are as precise as suggested in Table 4. Nevertheless, we believe this exercise helps demonstrate the utility of thinking about policies as elements of a "portfolio" that determines the variability of outcomes of interest. This is, of course, exactly the idea that Brainard originated in his seminal analysis.[98]

Overall, we believe this analysis provides support for a policy of tariff reduction for sub-Saharan Africa, unless one has very strong priors that a growth model applied to the rest of the world does not apply to that region.

7. Conclusions and suggestions for future research

In this paper, we have attempted to exposit a perspective on policy evaluation that explicitly places such evaluations in a decision-theoretic context and which explicitly accounts for uncertainty about the structure or model that describes the economic environment under analysis. On the theoretical side, this approach indicates that many of the standard objects of econometric study, for example evaluations of the statistical significance of a policy variable, may not be appropriate guides to policy analysis. The approach is also shown to allow for the evaluation of questions such as the robustness of policies in the presence of model uncertainty. We have also made some suggestions about how to implement this approach empirically. An example of empirical implementation to growth econometrics provided some additional insights relative to what is learned from more conventional approaches, although there are also important respects in which our new approach did not provide particularly different insights from what one finds from OLS exercises.
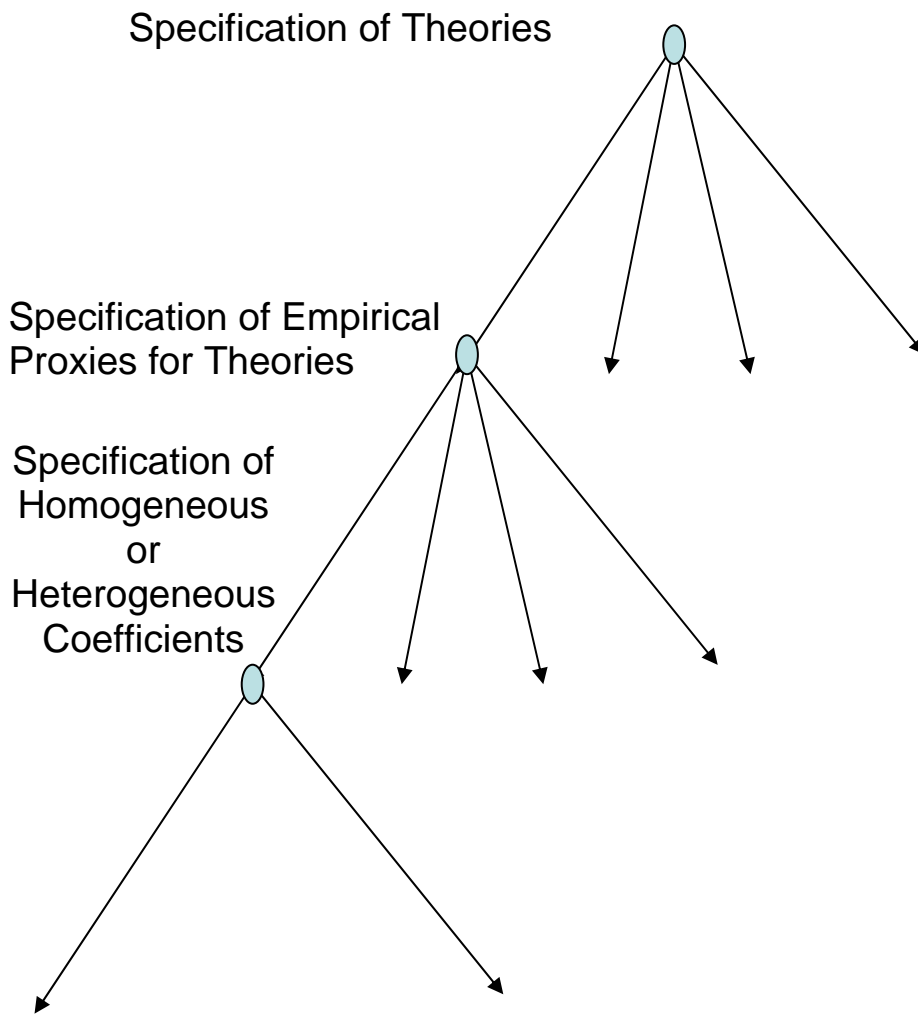
We reiterate that the methods we have described and the new literature in which it is situated still have far to go in terms of new methodological work. One important class of extensions may be defined in terms of generalizing our basic framework to better account for dynamics. For example, we have not dealt with issues relating to the

evolution of the model space. The averaging procedures we have described treat the model space as fixed; the only thing that evolves over time is the set of posterior model probabilities. This approach fails to incorporate the possibility that the set which a policymaker perceives as possible descriptions of the economy evolves over time; as we have argued earlier, this evolution has implications for whether the true model lies in the model space or not. Similarly, our analysis has not explicitly considered issues of policy choice when choices are updated across time in response to learning by the policymaker. Further, once learning is introduced, one can imagine an experimental design component to policy choice. A second important class of extensions concerns statistical issues. For example, our pseudo-Bayesian approach to integrating model uncertainty into a frequentist framework leads to a host of econometric questions in terms of how to do statistical inference for comparing the performance of different policy rules. Yet another question concerns possible nonlinearities in dynamic models; a body of work initiated by James Hamilton[99] suggests that the macroeconomy exhibits shifts across regimes; allowing for this possibility could prove to produce first-order effects in comparing stabilization policies. Regime shifts represent an additional layer of model uncertainty if a policymaker is not sure which regime is in effect when making a policy choice. Work is needed both to illustrate how to calculate policy effects accounting for possible nonlinearities (one loses the simple variance calculations that may be done with linear time series) as well as on the specification of model spaces and prior probabilities.

These limitations are not surprising, since the incorporation of model uncertainty into econometric analysis is still in its infancy. We believe that explicit attention to model uncertainty and the use of decision-theoretic methods will prove to be a fruitful direction for future macroeconomic research. At a minimum, explicitly accounting for model uncertainty in a decision-theoretic framework is an important step in clarifying the limits to which econometric analysis can contribute to policy evaluation.

Figure 1

STRUCTURE OF GROWTH MODELS

Specification of Theories

Specification of Empirical
Proxies for Theories

Specification of
Homogeneous
or
Heterogeneous
Coefficients

Table 1
Model Average Risk, for Optimal Rudebusch-Svensson Rules

λ_y	λ_i	g_π	g_y	g_i	Percentage increase risk, model averaging vs. Rudebusch-Svensson
0.00	0.10	4.5	2.0	0.2	306
0.00	0.50	2.3	1.0	0.4	17
0.00	1.00	1.7	0.7	0.5	-2
0.50	0.10	4.4	2.7	0.0	56
0.50	0.50	2.4	1.3	0.3	1
0.50	1.00	1.8	0.9	0.4	-9
1.00	0.10	4.3	3.2	-0.1	16
1.00	0.50	2.5	1.6	0.2	-7
1.00	1.00	1.7	1.0	0.4	-10
2.00	0.10	4.1	3.7	-0.2	8
2.00	0.50	2.5	1.9	0.1	-13
2.00	1.00	1.8	1.3	0.3	-14

Source: Authors' calculations.

Columns (λ_y) and (λ_i) report the assumed weights in the risk function (37),

$$R = var\left(\pi_t\right) + \lambda_y var\left(y_t\right) + \lambda_i var\left(\Delta i_t\right).$$

Columns (g_π), (g_y) and (g_i) report the optimal values for the monetary policy rule eq. (38),

$$i_t = g_\pi \pi_t + g_y y_t + g_i i_{t-1},$$

when the Rudebusch-Svensson (1999) model given by (40)-(41),

$$y_t = \alpha_{y1} y_{t-1} + \alpha_{y2} y_{t-2} - \alpha_{r1}(\bar{i}_{t-1} - \bar{\pi}_{t-1}) + u_t$$

$$\pi_t = \beta_{\pi1}\pi_{t-1} + \beta_{\pi2}\pi_{t-2} + \beta_{\pi3}\pi_{t-3} + \beta_{\pi4}\pi_{t-4} + \beta_{y1}y_{t-1} + v_t$$

is assumed to generate the data. The last column reports the percentage increase in risk when the model average risk (37) was used rather than the estimated risk using Rudebusch-Svensson, i.e.

$$100\times\left(\left(\frac{ER}{\hat{R}_{RS}}\right)-1\right).$$

Table 2
Tariff Parameter Evaluation Under Different Inclusion Probability (*q*)
Specifications

Model Specification	OLS	Full OLS	BMA			Min Coefficient	Max Coefficient	Min Mean + 2σ (Coefficient Plus Twice Std. Errors)	Max Mean + 2σ (Coefficient Plus Twice Std. Errors)
			q = .25	*q* = .50	*q* = .75				
Mean (standard deviation)	-.5377 (.2282)	-.5725 (.1977)	-.6022 (.1920)	-.5992 (.1885)	-.5946 (.1870)	-.6508 (.2163)	-.3595 (.2222)	-.6332 (.1898)	-.3737 (.2297)

Source: Authors' calculations.

The tariff variable measures tariffs on intermediate goods and inputs and corresponds to OWTI in the standard Barro and Lee data set.

Column (OLS) reports the OLS estimates based on (44),

$$g_i = \beta' X_i + \gamma' Z_i + \delta p_i + \varepsilon_i$$

for Solow variables and tariff.

Column (Full OLS) reports the OLS estimates based on (42) for the set of all available variables.

Columns (BMA) report the Bayesian Model Averaging estimates over alternative specifications of (44) with priors generated with inclusion probabilities, *q,* of .25, .50 and .75.

Column (Min Coefficient) reports the coefficient estimate and standard error for the version of (44) with the smallest coefficient estimate out of all the models used in the Bayesian Model Averaging exercises (variables included: Int, MNGD, MINV, MSCH, MGDP60, OWTI, RERD,GGCFD, CIVLIBb, DMCYBL).

Column (Max Coefficient) reports the coefficient estimate and standard error for the version of (44) with the largest coefficient estimate out of all the models used in the Bayesian Model Averaging exercises (variables included: Int, MNGD, MINV, MSCH, MGDP60, OWTI, BMPL, LLY, EcOrg, RULELAW, CIVLIBb).

Column (Min Mean $+ 2\sigma$) reports the coefficient estimate and standard error for the version of (44) where the coefficient estimate plus twice the standard error is smallest for all the models used in the Bayesian Model Averaging exercises (variables included: Int, MNGD, MINV, MSCH, MGDP60, OWTI, RERD, GVXDXE5, DMCYBL, PI6089, PIHYP6089).

Column (Max Mean $+ 2\sigma$) reports the coefficient estimate and standard error for the version of (44) where the coefficient estimate plus twice the standard error is largest for all the models used in the Bayesian Model Averaging exercises (variables included: Int, MNGD, MINV, MSCH, MGDP60, OWTI, DCPY, LLY, EcOrg, RULELAW, CIVLIBb).

Table 3
Tariff Coefficient Estimates Under Different Prior Heterogeneity Probabilities for Sub-Saharan Africa (SSA) for Theory Inclusion Probability q=0.5

Model Specification	OLS	Full OLS	BMA Prior Heterogeneity Probabilities		
			.5	*.75*	*1.00*
SSA	-.4320 (.8943)	-.2512 (1.0112)	-.6079 (.2205)	-.6246 (.2707)	-1.2322 (.7678)
Rest of World	-.6276 (.2067)	-.4630 (.2005)	-.5981 (.1890)	-.5961 (.1899)	-.5222 (.2067)

Source: Authors' calculations.

Column (OLS) reports the OLS estimates based on (46),

$$g_i = \beta' X_i + \gamma' Z_i + \delta p_i + \overline{\beta}' X_i \xi_{i,SSA} + \overline{\gamma}' Z_i \xi_{i,SSA} + \overline{\delta} p_i \xi_{i,SSA} + \varepsilon_i,$$

for Solow variables and tariff. Column (Full OLS) reports the OLS estimates based on (46) for the set of all available variables. Column (BMA Prior Heterogeneity Probabilities) reports the Bayesian Model Averaging estimates using versions of (46) with priors generated with an inclusion probability of .50 and where the prior probabilities of coefficient heterogeneity are respectively .5, .75, and 1.00.

Table 4: Expected Growth Change for SSA Countries Predicted by the Model as The Policy Variable (Tariff) is Decreased by 10%.
Heterogeneity for SSA is not Considered.
Inclusion Probability: .50

SSA Countries	*True Value*	Expected Growth Before Experiment	Expected Growth After Experiment	Change	Expected Growth from model with smallest coefficient on tariff			Expected Growth from model with biggest coefficient on tariff		
					Before Experiment	After Experiment	Change	Before Experiment	After Experiment	Change
Benin	-.0412	.1320 (.3438)	.1478 (.3439)	.0158 (.00002)	.2681 (.3461)	.2853 (.3459)	.0171 (-.0002)	.2728 (.3548)	.2823 (.3547)	.0094 (.0001)
Botswana	1.3423	1.0215 (.3437)	1.0344 (.3438)	.0129 (.00006)	1.0361 (.3592)	1.0501 (.3595)	.0140 (.0003)	1.0268 (.3586)	1.0345 (.3588)	.0077 (.0002)
Burkina Faso	.4824	.5338 (.3556)	.5627 (.3546)	.0288 (-.0010)	.3855 (.3699)	.4168 (.3685)	.0313 (-.0014)	.1983 (.3689)	.2157 (.3675)	.0173 (-.0014)
Burundi	-.1299	-.0898 (.3486)	-.0766 (.3486)	.0132 (-.00003)	-.3073 (.3596)	-.2929 (.3596)	.0143 (.00005)	-.1940 (.3608)	-.1861 (.3608)	.0079 (-.00005)
Cameroon	.9016	.5033 (.3399)	.5189 (.3398)	.0156 (-.0001)	.3448 (.3462)	.3617 (.3461)	.0169 (-.0001)	.7529 (.3529)	.7623 (.3527)	.0093 (-.0001)
Central Afr. Rep.	-.0603	.1417 (.3381)	.1536 (.3381)	.0119 (.00007)	.2866 (.3458)	.2996 (.3458)	.0130 (.00001)	.4198 (.3485)	.4270 (.3486)	.0071 (-.00005)
Congo	.9557	.6035 (.3524)	.6154 (.3525)	.0118 (.0001)	.7964 (.3559)	.8092 (.3559)	.0128 (-.00005)	.9217 (.3578)	.9289 (.3580)	.0071 (.0002)
Ethiopia	.1317	.2459 (.3452)	.2578 (.3455)	.0119 (.0002)	.2210 (.3528)	.2340 (.3531)	.0130 (.0003)	.1110 (.3602)	.1182 (.3603)	.0071 (.0001)
Gabon	1.4094	.8484 (.3400)	.8613 (.3399)	.0129 (-.00008)	.6303 (.3526)	.6443 (.3525)	.0140 (-.0001)	.7405 (.3570)	.7482 (.3570)	.0077 (-.00004)
Ghana	-.3278	-.1747 (.3558)	-.1550 (.3553)	.0197 (-.0004)	-.1407 (.3642)	-.1192 (.3637)	.0214 (-.0005)	.1138 (.3568)	.1256 (.3566)	.0118 (-.0001)
Ivory Coast	.2066	.3445 (.3457)	.3574 (.3456)	.0129 (-.00002)	.0845 (.3485)	.0985 (.3486)	.0140 (.0001)	.5129 (.3547)	.5207 (.3547)	.0077 (-.00001)
Kenya	.3421	.7192 (.3391)	.7357 (.3391)	.0164 (-.00001)	.6200 (.3467)	.6379 (.3467)	.0178 (-.00005)	.7114 (.3498)	.7213 (.3497)	.0098 (-.0001)
Madagascar	-.2026	.1932 (.3402)	.2085 (.3401)	.0152 (-.00004)	.1031 (.3457)	.1197 (.3457)	.0165 (.00003)	.2426 (.3592)	.2517 (.3591)	.0091 (-.0001)
Malawi	.5927	.8230 (.3513)	.8303 (.3515)	.0072 (.0002)	.7242 (.3564)	.7320 (.3566)	.0078 (.0002)	.5100 (.3597)	.5144 (.3599)	.0043 (.0002)
Mali	-.0373	.2486 (.3381)	.2615 (.3383)	.0129 (.0001)	.1460 (.3487)	.1600 (.3487)	.0140 (.00006)	.1680 (.3528)	.1757 (.3528)	.0077 (.00005)
Mauritania	.2896	.5856 (.3411)	.5985 (.3412)	.0129 (.00002)	.6351 (.3542)	.6491 (.3542)	.0140 (-.00005)	.6048 (.3558)	.6126 (.3559)	.0077 (.0001)
Mauritius	.4080	.5776 (.3358)	.5984 (.3353)	.0207 (-.0004)	.5172 (.3465)	.5398 (.3460)	.0225 (-.0005)	.6561 (.3495)	.6686 (.3490)	.0124 (-.0005)
Niger	.4449	.2695 (.3470)	.2825 (.3470)	.0129 (-.00002)	.0921 (.3517)	.1061 (.3518)	.0140 (.00006)	.4615 (.3604)	.4693 (.3604)	.0077 (-.00006)
Nigeria	.1170	-.1091 (.3563)	-.0823 (.3551)	.0267 (-.0011)	-.2440 (.3681)	-.2149 (.3668)	.0290 (-.0013)	.1940 (.3493)	.2101 (.3483)	.0160 (-.0010)
Rwanda	.4141	.1590 (.3404)	.1755 (.3404)	.0164 (.00004)	.0681 (.3534)	.0860 (.3534)	.0178 (.00003)	.0223 (.3559)	.0321 (.3559)	.0098 (-.0001)
Senegal	.0408	.0983 (.3347)	.1096 (.3347)	.0113 (.00006)	.1258 (.3441)	.1381 (.3442)	.0122 (.00007)	.1546 (.3498)	.1614 (.3499)	.0068 (.0001)
Sierra Leone	.4545	.0947 (.3549)	.1020 (.3550)	.0073 (.0001)	.4270 (.3505)	.4350 (.3507)	.0079 (.0002)	.5964 (.3588)	.6008 (.3590)	.0043 (.0002)
Somalia	-.3158	.1451 (.3466)	.1573 (.3466)	.0122 (.00005)	.2648 (.3524)	.2781 (.3525)	.0132 (.00005)	.2932 (.3511)	.3005 (.3512)	.0073 (.0001)
South Africa	.3931	.4022 (.3382)	.4151 (.3381)	.0129 (-.0001)	.4877 (.3589)	.5018 (.3588)	.0140 (-.0001)	.3856 (.3555)	.3934 (.3550)	.0077 (-.0005)

64

Sudan	-.1890	-.0419 (.3417)	-.0221 (.3414)	.0198 (-.0003)	.1926 (.3504)	.2142 (.3499)	.0215 (-.0004)	.1922 (.3484)	.2041 (.3480)	.0119 (-.0004)
Tanzania	.6172	.5078 (.3515)	.5181 (.3517)	.0103 (.0002)	.5549 (.3594)	.5661 (.3596)	.0111 (.0002)	.4754 (.3674)	.4816 (.3676)	.0061 (.0002)
Togo	.2301	.6298 (.3372)	.6427 (.3374)	.0129 (.0001)	.6659 (.3462)	.6799 (.3463)	.0140 (.00008)	.7188 (.3508)	.7265 (.3510)	.0077 (.0002)
Uganda	.1042	-.3093 (.3491)	-.3031 (.3493)	.0061 (.0001)	-.1412 (.3568)	-.1345 (.3571)	.0067 (.0003)	-.2827 (.3678)	-.2790 (.3681)	.0037 (.0003)
Zaire	-.3659	.0333 (.3487)	.0406 (.3489)	.0073 (.0001)	.1260 (.3599)	.1339 (.3602)	.0079 (.0003)	.3452 (.3565)	.3496 (.3568)	.0043 (.0003)
Zambia	-.1472	.2496 (.3575)	.2605 (.3576)	.0109 (.00003)	.6691 (.3588)	.6810 (.3589)	.0119 (.0001)	.6025 (.3578)	.6091 (.3579)	.0065 (.0001)
Zimbabwe	.5738	.8728 (.3374)	.8865 (.3374)	.0137 (.00005)	.6666 (.3462)	.6815 (.3462)	.0149 (.00007)	.6326 (.3465)	.6408 (.3467)	.0082 (.0002)

Source: Authors' calculations.

Table 4 reports an exercise for the sub-Saharan African economies in which the mean and variance of the growth rate for each country between 1960 and 1985 is compared with and without a 10% reduction of tariffs as compared to what occurred historically. To do this, we use the posterior means and variances of the model parameters β, γ, and δ based on the historical data. We then compute the posterior mean and variance of g_i with and without a 10% reduction in the tariff variable, keeping all other regressor values constant. We assume that the errors in the growth process are independent of the regressors. Standard deviations are shown in parentheses.

Computational Appendix

Posterior Coefficient Densities

Posterior densities for the parameters of growth models were calculated under the following assumptions. For a given regression, let S_i denote the regressor associated with country i. A growth regression will therefore have the form

$$g_i = S_i \zeta + \varepsilon_i \quad i = 1 \ldots I \tag{47}$$

In order to compute the posterior distribution of ζ given data and a specific model, i.e. $\mu(\zeta|d,m)$, we assume first that there is no informative prior information available on the coefficients. In more standard language, we impose a noninformative prior on the coefficients, i.e.

$$\mu(\zeta) \propto c \tag{48}$$

Second, we assume that the errors are i.i.d. normal with a known variance. Under this assumption, one can show[100] that the posterior density of the regression coefficients is

$$\mu(\zeta|d,m) \square N\left(\hat{\zeta}, (S'S)^{-1}\sigma_\varepsilon^2\right) \tag{49}$$

where $\hat{\zeta}$ is the OLS estimate of the coefficient parameters in (47). Notice also that $(S'S)^{-1}\sigma_\varepsilon^2$ is the OLS variance estimate for the parameters when the error variance is known. A helpful feature of this formula is that it means that the parameters of the posterior density of ζ have OLS interpretations. The assumption that the error variance is known is not serious when the number of observations is large relative to the number of regressors.

We should note that there is a considerable discussion in the literature about the appropriate choice of priors even for this model. Fernandez, Ley and Steel[101] consider a range of alternative priors and argue in favor of a different set of priors than those we employ. We do not claim that our choice of priors is in any sense optimal; we employ it here in order to produce a close relationship between OLS estimates and Bayesian posterior estimates.

Model Averaging Calculations

1. Monetary policy

All Bayesian model averaging exercises in the monetary policy section of the paper were performed using RATS.

2. Growth

All Bayesian model averaging exercises in the growth section of the paper were calculated using SPLUS. The number of models under study was small enough to allow the analysis to calculate posterior coefficient densities using all available models. For larger exercises, it is necessary to use a search algorithm to focus on models with relatively large posterior probabilities. One such program is *bicreg* written by Adrian Raftery and available at www.research.att.com/~volinsky/bma.html. This procedure uses an "Occam's Window" procedure due to David Madigan and Raftery.[102] In adapting the code for our exercise, a few adjustments were necessary that are available from the authors.

Prior probabilities were set as follows. For a given growth specification, one first specifies the probability a given theory is included. Table 2 allows these probabilities to be .25, .5, and .75. For a given theory, with r empirical proxies, there are $2^r - 1$ different ways to include these proxies. Each specification is assumed to have equal ex ante probability. Table 3 reports results where each specification of a set of theories and empirical proxies used to calculate Table 2 is matched with a corresponding model with

sub-Saharan Africa heterogeneity, with corresponding specifications given equal probability.

The calculation of posterior model probabilities can also be computationally difficult. In order to handle these calculations, we follow an approximation suggested by Raftery[103] which exploits the fact if the data under study fulfill the necessary conditions for posterior coefficient distributions to converge to their associated maximum likelihood estimators, one can use the maximum likelihood estimates as approximations to the posterior distributions and therefore avoid the need to specify a particular prior on the coefficients within a model; in essence the weights are BIC adjusted likelihoods. This greatly simplifies the calculations of posterior model probabilities.[104] Of course, the approximation becomes more accurate the larger the data set. The program for this approximation is taken from *bicreg* described above.

Data Appendix

1. Monetary policy

In the monetary policy section, all data were obtained from the Federal Reserve Bank of St. Louis website. Real GDP is measured in chained 1996 dollars, with inflation measured by the corresponding price index. Potential GDP is the Congressional Budget Office measure. The quarterly average Federal Funds rate was computed by averaging monthly average figures.

2. Growth

The various growth variables were taken from a range of sources.

Solow variables

MNGD: $\ln(n + g + d)$ where n =population growth, g =exogenous rate of technical change, and d =depreciation. $g + d$ is assumed to equal .05 for all countries. Source: Mankiw, Romer, Weil (1992)

MINV: log of the investment rate. Source, Mankiw, Romer, Weil (1992)

ln(SCHOOL); log of the fraction of the population between ages 12 and 17 enrolled in school multiplied by fraction of working age population between ages of 15 and 19. Source: Mankiw, Romer and Weil (1992).

MGDP60, log of per capita income in 1960, Source: Mankiw, Romer and Weil (1992).

Note: a constant term is always included as a Solow regressor.

Policy variables

1. Tariffs

OWTI: Own Import Weighted Tariff Rates on Intermediate Inputs and Capital Goods; Source: Barro and Lee (1994).

2. Exchange rates

BMPL6089: Black Market Premium. 30-year averages: 1960 - 1989. log(1+BMP). Source: Barro and Lee (1994).

RERD: Real Exchange Rate Distortions; Source: Dollar (1992).

3. Inflation

PI6089: Average Inflation Rate for the period 1960-89; Sala-i-Martin (1997). Original Source: Levine and Renelt (1992).

PIHYP6089: Dummy for Average Inflation Rate for period 1960-89 above 15%. Created from PI6089.

4. Government spending

GGCFD: Ratio of Real Public Domestic Investment to Real GDP; Barro and Lee (1994).

GVXDXE5: Ratio of Real Government "Consumption" Expenditure Net of Spending on Defense and on Education to Real GDP; Barro and Lee (1994).

Structural variables

1. Economic structure

EcOrg: Capitalism. Index of degree of capitalism as measured by Freedom House (1994). Source: Sala-i-Martin (1997).

RULELAW: Index of rule of law; Source: Sala-i-Martin (1997).

2. Financial structure

DCPY: Ratio of Gross Claims on the Non-Financial Private Sector by Central Bank and Deposit Banks to GDP; Source: King and Levine (1993)

LLY: Ratio of Liquid Liabilities of the financial system to GDP; Source: King and Levine (1993)

3. Political structure

Civilly: Index of civil liberties. Source: Knack and Keefer (1995).

DMCYBL: Democracy Index. Index from 0 to 1; 1=most democratic. Source: Barro and Lee (1994).

Bibliography

Azariadis, Costas and Allan Drazen. 1990. "Threshold Externalities in Economic Development." *Quarterly Journal of Economics* 105: 501-26.

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

Barro, Robert and Jong-Wha Lee. 1994. "Data Set for a Panel of 134 Countries." *National Bureau of Economic Research*, available at www.nber.org/pub/barro.lee.

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

Becker, Selwyn and Fred Brownson. 1964. "What Price Ambiguity? Or the Role of Ambiguity in Decisionmaking." *Journal of Political Economy* 72: 62-73.

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

Bernard, Andrew and Steven Durlauf. 1996. "Interpreting Tests of the Convergence Hypothesis." *Journal of Econometrics* 71: 161-73.

Bernardo, Jose and Adrian Smith. 1994. *Bayesian Theory*, New York: John Wiley.

Bernhard, Pierre. 2002. "Survey of Linear Quadratic Robust Control." *Macroeconomic Dynamics* 6: 19-39.

Blinder, Alan. 1997. "What Central Bankers Could Learn From Academics and Visa Versa." *Journal of Economic Perspectives* 11(2): 3-19.

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

Brainard, William. 1967. "Uncertainty and the Effectiveness of Policy." *American Economic Review* 57: 411-25.

Brock, William and Steven Durlauf. 2001. "Growth Economics and Reality." *World Bank Economic Review* 15(2): 229-72.

------. 2003. "Design Limits Analysis and Local Robustification Analysis: Theory and Applications to Economics," in progress.

Brown, Philip, Marina Vannucci, and Tom Fearn. 1998. "Multivariate Bayesian Variable Selection and Prediction." *Journal of the Royal Statistical Society, series B* 60: 627-41.

------. 2002. "Bayes Model Averaging with Selection of Regressors." *Journal of the Royal Statistical Society, series B* 64: 519-36.

Camerer, Colin. 1995. "Individual Decisionmaking." In *Handbook of Experimental Economics*, edited by John Kagel and Alvin Roth. Princeton: Princeton University Press.

Canova, Fabio. 1999. "Testing for Convergence Clubs in Cross-Country Growth Data: A Predictive Density Approach." Unpublished paper. Department of Economics, University of Pompeu Fabra, Spain.

Chamberlain, Gary. 2001. "Econometrics and Decision Theory." *Journal of Econometrics* 95: 255-83.

Chipman, Hugh, Edward George, and Robert McCulloch. 2001. "The Practical Implementation of Bayesian Model Selection." In *Model Selection*, edited by P. Lahiri. Hayward: Institute of Mathematical Statistics Lecture Notes-Monograph Series vol. 38.

Cox, David and David Hinkley. 1974. *Theoretical Statistics*. New York: Chapman and Hall.

Curley, Shawn, Frank Yates and Richard Abrams. 1986. "Psychological Sources of Ambiguity Avoidance." *Organization Behavior and Human Decision Processes* 38: 230-56.

Dehejia, Rajeev. 2001. "Program Evaluation as a Decision Problem." Unpublished paper. Columbia University. Forthcoming, *Journal of Econometrics.*

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

Dollar, David. 1992. "Outward-Oriented Developing Economies Really Do Grow More Rapidly: Evidence from 95 LDC's." *Economic Development and Cultural Change* 40: 523-44.

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

Draper, David. 1995. "Assessment and Propagation of Model Uncertainty." *Journal of the Royal Statistical Society, series B* 57: 45-70.

Durlauf, Steven. 2000. "Econometric Analysis and the Study of Economic Growth: A Skeptical Perspective." In *Macroeconomics and the Real World*, edited by Roger Backhouse and Andrea Salanti. Oxford: Oxford University Press.

Durlauf, Steven and Paul Johnson. 1995. "Multiple Regimes and Cross-Country Growth Behavior." *Journal of Applied Econometrics* 10(4): 363-84.

Durlauf, Steven, Andros Kourtellos and Artur Minkin. 2001. "The Local Solow Growth Model." *European Economic Review* 45: 928-40.

Durlauf, Steven. and Danny Quah. 1999. "The New Empirics of Economic Growth." In *Handbook of Macroeconomics*, edited by John Taylor and Michael Woodford. Amsterdam: North Holland.

Easterly, William and Ross Levine. 1997. "Africa's Growth Tragedy: Politics and Ethnic Divisions." *Quarterly Journal of Economics* 112: 1203-50.

Ellsberg, Daniel. 1961. "Risk, Ambiguity, and the Savage Axioms." *Quarterly Journal of Economics* 75(4): 643-69.

Epstein, Larry and Tau Wang. 1994. "Intertemporal Asset Pricing Behavior Under Knightian Uncertainty." *Econometrica* 62: 283-322.

Fernandez, Carmen, Eduardo Ley and Mark Steel. 2001a. "Benchmark Priors for Bayesian Model Averaging." *Journal of Econometrics* 100(2): 381-427.

------. 2001b. "Model Uncertainty in Cross-Country Growth Regressions." *Journal of Applied Econometrics* 16(5): 563-76.

French, Simon and David Rios Insua. 2000. *Statistical Decision Theory*. London: Arnold.

George, Edward. 1999. "Discussion of "Bayesian Model Averaging and Model Search Strategies" by M. Clyde." In *Bayesian Statistics 6*, edited by Jose Bernardo and others. Oxford: Oxford University Press.

Giannoni, Marc. 2002. "Does Model Uncertainty Justify Caution? Robust Optimal Monetary Policy in a Forward-Looking Model." *Macroeconomic Dynamics* 6: 111-144.

Gilboa, Itzhak and David Schmeidler. 1989. "Maximin Expected Utility with Non-Unique Priors." *Journal of Mathematical Economics* 18: 141-53.

Hamilton, James. 1989. "A New Approach to the Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57: 357-84.

Hansen, Lars and Thomas Sargent. 2001a. "Acknowledging Misspecification in Macroeconomic Theory." *Review of Economic Dynamics* 4: 519-35.

------. 2001b. "Robust Control and Model Uncertainty." Unpublished paper. Hoover Institution, Stanford University.

------. 2002a. *Robust Control and Economic Model Uncertainty*. Book manuscript. Hoover Institution, Stanford University.

------. 2002b. ""Certainty Equivalence" and "Model Uncertainty"." Unpublished paper. Hoover Institution, Stanford University.

von Hayek, Friedrich. 1942. "Scientism and the Study of Society." *Economica* 9(35): 267-91.

Heckman, J. 2001a. "Accounting for Heterogeneity, Diversity and General Equilibrium in Evaluating Social Programmes," *Economic Journal*, 111: F654-F699.

-----. 2001b. "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture," *Journal of Political Economy*, 109: 673-748.

Hoeting, Jennifer and others. 1999. "Bayesian Model Averaging: A Tutorial." *Statistical Science* 14: 382-401.

Hurwicz, Leonid. 1951. "Some Specification Problems and Applications to Econometric Models." *Econometrica* 19: 343-4.

Kadane, Joseph, Ngai Hang Chan and Lara Wolfson. 1996. "Priors for Unit Root Models." *Journal of Econometrics* 75: 99-111.

Keynes, John Maynard. 1940. "On a Method of Business-Cycle Research: A Comment." *Economic Journal* 50(197): 154-6.

King, Roland and Ross Levine. 1993. "Finance and Growth: Schumpeter Might Be Right." *Quarterly Journal of Economics* 108(3): 717-37.

Knack, Stephen and Philip Keefer. 1995. "Institutions and Economic Performance: Cross-Country Tests Using Alternate Institutional Measures." *Economics and Politics* 7(3): 207-27.

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

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

Leamer, Edward and Herman Leonard. 1983. "Reporting the Fragility of Regression Estimates." *Review of Economics and Statistics* LXV(2): 306-17.

Leeper, Eric, Christopher Sims and Tau Zha. 1996. "What Does Monetary Policy Do? (with discussion)." *BPEA*, 2, 1-78.

Levin, Andrew and John Williams. 2002. "Robust Monetary Policy with Competing Reference Models." *Journal of Monetary Economics* (forthcoming).

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

Lindley, David. 1990. "The 1988 Wald Memorial Lectures: The Present Position in Bayesian Statistics (with discussion)." *Statistical Science* 5(1): 44-89.

Madigan, David and Adrian Raftery. 1994. "Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam's Window." *Journal of the American Statistical Association* 89: 1535-46.

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

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

Manski, Charles. 2001. "Identification Problems and Decisions Under Ambiguity; Empirical Analysis of Treatment Response and Normative Analysis of Treatment Choice." *Journal of Econometrics* 95(2): 415-42.

------. 2002. "Treatment Choice Under Ambiguity Induced by Inferential Problems." *Journal of Statistical Planning and Inference* 105(1): 67-82.

------. 2003. "Social Learning from Private Experiences: The Dynamics of the Selection Problem." *Review of Economic Studies*, forthcoming.

Maritz, J. and T. Lwin. 1989. *Empirical Bayes Methods*, *second edition*. London: Chapman and Hall.

McAleer, Michael, Adrian Pagan, and Paul Volker. 1983. "What Will Take the Con Out of Econometrics?" *American Economic Review* 75: 293-307.

McCallum, Bennett. 1988. "Robustness Properties of a Rule for Monetary Policy." *Carnegie Rochester Conference Series on Public Policy* 29: 173-203.

McConnell, Margaret and Gabriel Perez-Quiros. 2000. "Output Fluctuations in the United States: What Has Changed since the Early 1980's?" *American Economic Review* 90: 1464-76.

Onatski, Alexei. and James Stock. 2002. "Robust Monetary Policy Under Model Uncertainty in a Small Model of the U.S. Economy." *Macroeconomic Dynamics* 6: 85-110.

Onatski, Alexei. and Noah Williams. 2003. "Modeling Model Uncertainty." Unpublished paper. Princeton University.

Phillips, Peter. 1991. "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends." *Journal of Applied Econometrics* 6: 333-64.

Raftery, Adrian. 1995. "Bayesian Model Selection in Social Research (with discussion)." In *Sociological Methodology 1995,* edited by Peter V. Marsden. Cambridge, MA: Blackwell

Raftery, Adrian, David Madigan, and Jennifer Hoeting. 1997. "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association* 92(437): 179-91.

Raiffa, Howard and Robert Schlaifer. 1961. *Applied Statistical Decision Theory*. Reprint. New York: John Wiley and Sons Classic Library, 2000.

Rubin, Donald. 1980. "Using Empirical Bayes Techniques in Law School Validity Studies." *Journal of the American Statistical Association* 75: 801-816.

Rudebusch, Glenn and Lars Svensson. 1999. "Policy Rules for Inflation Targeting." In *Monetary Policy Rules*, edited by John Taylor. Chicago: University of Chicago Press.

Sala-i-Martin, Xavier. 1997. "I Just Ran 2 Million Regressions." *American Economic Review* 87: 178-83.

Sims, Christopher. 1988. "Bayesian Skepticism on Unit Root Econometrics." *Journal of Economic Dynamics and Control* 12: 463-474.

------. 2001. "Pitfalls of a Minimax Approach to Model Uncertainty." *American Economic Review* 91(2): 51-4.

------. 2002. "The Role of Models and Probabilities in the Monetary Policy Process." *BPEA,* 2, 1-40.

Slovic, Paul and Amos Tversky. 1974. "Who Accepts Savage's Axioms?" *Behavioral Science* 19: 368-73.

Svensson, Lars. 1996. "Commentary: How Should Monetary Policy Respond to Shocks While Maintaining Long Run Price Stability?" In *Achieving Price Stability*, Federal Reserve Bank of Kansas City, 181-195.

Tan, Chih Ming. 2003. "Geography, Institutions, and Diversity as Factors Underlying Heterogeneity: A Bayesian Treed Regression Approach to Cross-Country Differences in Economic Performance." Unpublished paper. University of Wisconsin

Taylor, John. 1996. "How Should Monetary Policy Respond to Shocks While Maintaining Long Run Price Stability?" In *Achieving Price Stability*, Federal Reserve Bank of Kansas City.

------. 1999a. "The Robustness and Efficiency of Monetary Policy Rules as Guidelines for Interest Rate Setting by the European Central Bank." *Journal of Monetary Economics* 43: 655-79.

------. (1999b), *Monetary Policy Rules*, Chicago: University of Chicago Press.

Temple, Jonathan. 2000. "Growth Regressions and What the Textbooks Don't Tell You." *Bulletin of Economic Research* 52: 181-205.

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

Wald, Abraham. 1950. *Statistical Decision Functions*. New York: John Wiley.

Wasserman, Larry. 2000. "Bayesian Model Selection and Model Averaging." *Journal of Mathematical Psychology* 44: 92-107.

Woodford, Michael. 2002. "Inflation Stabilization and Welfare." Unpublished manuscript. Princeton University.

Endnotes

---

[2] von Hayek (1942, p.290).

[3] Keynes (1940, pp. 155-156).

[4] Brainard (1967).

[5] Heckman (2001b) is a brilliant overview of this work.

[6] Chamberlain (2001).

[7] Sims (2002).

[8] Dehejia (2001) provides an example of how such an approach may be used in practice.

[9] Manski (2001,2002).

[10] cf. Leamer (1978,1983), Leamer and Leonard (1983).

[11] Important examples include Draper (1995), Raftery, Madigan, and Hoeting (1997) and Chipman, George, and McCullough (2001).

[12] Hansen and Sargent (2001a,b,2002a,b).

[13] Levin and Williams (2002) is a recent contribution that is related to the analysis reported here.

[14] Levine and Renelt (1992).

[15] Sala-i-Martin (1997).

[16] Doppelhofer, Miller, and Sala-i-Martin (2000); Fernandez, Ley and Steel (2001b).

[17] See also Brock and Durlauf (2001)).

[18] Good surveys include Berger (1987), French and Rios Insua (2000) and the classic Raiffa and Schlaifer (1961); Cox and Hinkley (1974) chapter 11 is a very well written introduction.

[19] Chamberlain (2001) is a nice recent example.

[20] For example, Taylor (1999a).

[21] Barro and Xavier Sala-i-Martin (1995).

[22] cf. Sims (2002).

[23]An important exception to this holds for unit roots in time series, which has led to considerable controversy on how to implement Bayesian methods and in turn on how to interpret frequentist analyses, see Sims (1988), Phillips (1991), and Kadane, Chan, and Wolfson (1996) for further discussion. Much of the disagreement one finds in this literature concerns the appropriateness of various priors that have been proposed for autoregressive models, specifically with respect to the prior probability placed on unit root or integrated models. Without taking a stance in this debate, we do observe that the

sensitivity of results to priors that occurs in unit roots is in our judgment a major reason why Bayesian methods have not become more widespread in this context.

[24] For example, one can consider hierarchical models in which the parameters of a model are themselves functions of various observables and unobservables. If these relationships are continuous, one can trace out a continuum of models.

[25] See Wasserman (2000) for a very clear introduction and Raftery, Madigan, and Hoeting (1997) and Hoeting, Madigan, Raftery, and Volinsky (1999) for a detailed development of the technique in the context of regression models.

[26] Levin and Williams (2002).

[27] Leamer (1978, p. 118).

[28] Draper (1995).

[29] See Camerer (1995) for a superb survey of experiments that challenge aspects of expected utility theory.

[30] Ellsberg (1961).

[31] Becker and Brownson (1964).

[32] Slovic and Tversky (1974).

[33] Curley, Yates, and Abrams (1986).

[34] Epstein and Wang (1994) and Gilboa and Schmeidler (1989).

[35] Epstein and Wang (1994).

[36] See Hurwicz (1951) for a remarkable early suggestion that preferences like (8) could bridge different variants of statistical decision theory; these types of preferences are discussed in Manski (2003).

[37] Brock and Durlauf (2001).

[38] McAleer, Pagan, and Volker (1983).

[39] Brock and Durlauf (2001).

[40] Lars Hansen and Thomas Sargent (2001a,b,2002a,b)

[41] Examples include Giannoni (2002), Onatksi and Stock (2002).

[42] Wald (1950).

[43] Hansen and Sargent (2001a, p. 523).

[44] Sims (2001) critiques the use of robust control theory for failing to accommodate large model uncertainty in the sense of considering model uncertainty when there are multiple core models.

[45] Taylor (1999a, p. 658).

[46] Additional development of the specific approach we take may be found in Brock and Durlauf (2003).

[47] A more detailed analysis may be found in Brock and Durlauf (2003).

[48] Brainard (1967).

[49] Svensson (1996).

[50] Rudebusch and Svensson (1999).

[51] See Bernhard (2002) for an extended analysis which makes this intuition formal and Giannoni (2002) for a range of interesting findings along these lines**.**

[52] Taylor (1996, p. 185).

[53] Lindley (1990, p. 54).

[54] Leeper, Sims, and Zha (1996, p. 5).

[55] We say approximate since standard calculations of the likelihood statistic are calculations that substitute the maximum likelihood estimates of parameters for the model parameters. This type of substitution is in fact common in Bayesian calculations that approximate posterior model probabilities; see Raftery (1995), for example.

[56] Similarly, strict Bayesians object to empirical Bayes methods (Maritz and Lwin (1989)), but such techniques have proven useful in practice, despite alleged philosophical shortcomings, cf. Rubin (1980).

[57] Cox and Hinkley (1974, p. 429).

[58] Raiffa and Schlaifer (1961, p. 15).

[59] Bernardo and Smith (1994).

[60] Fernandez, Ley and Steel (2001a).

[61] This is the property defended by Leeper, Sims and Zha (1996) in their discussion of priors which we cited earlier.

[62] Berger (1987, ch. 4).

[63] Leeper, Sims, and Zha (1996).

[64] cf. Fernandez, Ley and Steel (2001a), Raftery, Madigan and Hoeting (1997).

[65] See Doppelhofer, Miller, and Sala-i-Martin (2000).

[66] Brown, Vannucci, and Fearn (1998,2002).

[67] Brock and Durlauf (2001).

[68] Rudebusch and Svensson (1999).

[69] McCallum (1988).

[70] Taylor (1999b); Levin and Williams (2002).

[71] Onatski and Stock (2002); Onatski and Williams (2003).

[72] See Woodford (2002).

[73] McConnell and Perez-Quiros (2000).

[74] See Onatski and Williams (2003) for a discussion of alternative treatments and interpretation of nonstationary estimates.

[75] Rudebusch and Svensson (1999).

[76] Levin and Williams (2002).

[77] Levin and Williams (2002).

[78] See Brock and Durlauf (2001), Durlauf (2000), and Temple (2000) for related analyses.

[79] See Barro (1991), Mankiw, Romer, and Weil (1992).

[80] Malinvaud (1998, p. 781).

[81] Barro and Sala-i-Martin (1995).

[82] Extended criticisms of cross-country growth regressions include Brock and Durlauf (2001) and Temple (2000). A number of these criticisms may be interpreted as arguing that standard growth analyses fail to properly account for model uncertainty.

[83] Durlauf and Quah (1999).

[84] Brock and Durlauf (2001).

[85] In addition, many modern growth theories imply that the growth process is fundamentally nonlinear. One important example of this type of model is due to Azariadis and Drazen (1990), who develop a model in which multiple steady states exist, with sufficiently poor economies subject to development traps. As shown in Durlauf and Johnson (1995), cross-country data generated by this model will have the properties that various subsets of economies will be associated with distinct linear models. These

distinct models identify countries that are associated with a common steady state. As a result, linear regressions will poorly approximate the growth process; see Bernard and Durlauf (1996) for discussion.

[86] See Canova (1999), Desdoigts (1999), Durlauf and Johnson (1995), Durlauf, Kourtellos and Minkin (2001) and Tan (2003).

[87] The empirical growth literature has become increasingly sensitive to the problem of parameter heterogeneity in the sense that it is now common to add dummy variables for certain regions or collections of countries, and occasionally to add interaction terms that multiply per capita income, say, with some growth determinant. It seems fair to say these efforts are generally ad hoc.

[88] Brock and Durlauf (2001).

[89] Blinder (1997, pg. 6) makes a similar criticism of quadratic loss functions for policy analysis in business cycle contexts. Our own view is that the limitations of quadratic loss functions for business cycle analysis largely stem from their failure to accommodate issues of distribution effects, and so that a proper development of alternative loss functions would simultaneously need to address the question of how to introduce the measurement of distribution effects in empirical business cycle analysis. See Heckman (2001a) for important recent work on policy evaluation when effects are heterogeneous.

[90] There are 31 countries in sub-Saharan Africa for which data are available across the time period we consider: Benin, Botswana, Burkina Faso, Burundi, Cameroon, Central African Republic, Congo, Ethiopia, Gabon, Ghana, Ivory Coast, Kenya, Madagascar, Malawi, Mali, Mauritania, Mauritius, Niger, Nigeria, Rwanda, Senegal, Sierre Leone, Somalia, South Africa, Sudan, Tanzania, Togo, Uganda, Zaire, Zambia, and Zimbabwe.

[91] This approach corresponds to Leamer's (1983) distinction between "maintained" and "doubtful" variables. In using this approach, we are conducting a different exercise from that done in Doppelhofer, Miller, and Sala-i-Martin (2000) or Fernandez, Ley, and Steel (2001b), where the focus is on identifying which variables should be included in a growth model from a large set of potential growth determinants and where no distinctions are made between the prior inclusion probabilities for different variables. We make such a distinction as we include the Solow variables and tariff variable with probability 1.

[92] George (1999) suggests a method for accounting for similarities in models by a method he refers to as "dilution priors." The approach does not appear to have yet been formalized in the statistics literature, but its logic appears similar to the tree structure approach we employ.

[93] Brock and Durlauf (2001); Easterly and Levine (1997).

[94] See Brock and Durlauf (2001), Doppelhofer, Miller, Sala-i-Martin (2000), Fernandez, Ley, and Steel (2001b).

[95] For example, we are currently exploring ways to treat observed control variables as proxies for underlying theories, so that the way they appear in a model conditional on the inclusion of a theory is handled in a way that the empirical proxies are combined to form an optimal estimate of the empirical "signal" associated with the theory.

[96] Fernandez, Ley and Steel (2001a).

[97] Brock and Durlauf (2001).

[98] Brainard (1967).

[99] Hamilton (1989).

---

[100] cf. Box and Tiao (1973, p. 115).

[101] Fernandez, Ley and Steel (2001a).

[102] Madigan and Raftery (1994).

[103] Raftery (1995).

[104] Raftery (1995) as well as Tierney and Kadane (1986) contain technical arguments to justify this approximation. Doppelhofer, Miller, and Sala-Martin (2000) use the same approach to averaging and provide a justification for using diffuse priors when comparing models, which can be a problematic issue in Bayesian contexts.