WILLIAM A. BROCK
University of Wisconsin

STEVEN N. DURLAUF
University of Wisconsin

KENNETH D. WEST
University of Wisconsin

Policy Evaluation in Uncertain Economic Environments

> It will be remembered that the seventy translators of the Septuagint were shut up in seventy separate rooms with the Hebrew text and brought out 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.[1]

THIS PAPER DESCRIBES SOME approaches to macroeconomic policy evaluation in the presence of uncertainty about the structure of the economic 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 to apply, or about the particular functional forms that translate a general theory into a form amenable to statistical analysis. As such, the methods we describe are, we believe, particularly useful in a range of macroeconomic contexts where fundamental disagreements exist as to the determinants of the problem under study. In addition, this approach recognizes that even if economists agree on the

We thank the John D. and Catherine T. MacArthur Foundation and the National Science Foundation, the Vilas Trust, and the University of Wisconsin Graduate School for financial support. This paper was written while Kenneth West was a visitor at Victoria University, Wellington, and the Reserve Bank of New Zealand, whose hospitality he gratefully acknowledges. We are especially grateful to Ritesh Banerjee, Ethan Cohen-Cole, Artur Minkin, Giacomo Rondina, Chih Ming Tan, and Yu Yuan for outstanding research assistance. Our discussants and the other conference participants provided valuable suggestions.

1. Keynes (1940, pp. 155–56).

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 the theory does not specify. Hence our analysis is motivated by concerns similar to those that led to the development of model calibration methods. Unlike in the usual calibration approach, however, 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 on the basis of two factors: a policymaker's preferences, and 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, namely, 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, a major endeavor in much empirical macroeconomic research, is not a necessary component of policy evaluation.

To the contrary: our argument is that, in many cases, model selection is actually inappropriate, because conditioning policy evaluation on a particular model ignores the role of model uncertainty in the overall uncertainty that surrounds 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 a theoretical perspective, model uncertainty has important implications for the evaluation of policies. This was originally recognized in William Brainard's classic analysis,[2] 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 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 gener-

2. Brainard (1967).

alization 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 integrating 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 a policy's efficacy 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 attempt to place policy theoretical and empirical evaluation exercises in a framework that properly accounts both for the decision-theoretic nature of the question and for the different types of uncertainty. In this we are motivated by concerns similar to those that have influenced a number of other researchers. Many of James Heckman's contributions may be interpreted as providing methods for policy analysis that properly account for the ways in which individuals make decisions.[3] Gary Chamberlain and Christopher Sims have argued explicitly in favor of Bayesian decision-theoretic approaches to data analysis.[4] Charles Manski has advocated, in contexts where one cannot identify which of several models explains a given data set, an approach that focuses on finding "undominated" policies, that is, policies that are optimal for at least one model consistent with the data.[5] 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 and where it 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.

3. Heckman (2001b) is a brilliant overview of this work.

4. Chamberlain (2001); Sims (2002). Dehejia (forthcoming) provides an example of how such an approach may be used in practice.

5. Manski (2001, 2002).

We are also far from being 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.[6] Leamer's ideas have motivated a number of recent developments in the statistics literature.[7] In terms of the theory of policy analysis, Lars Hansen and Thomas Sargent, among others, 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.[8] 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.

The problem of model uncertainty has also motivated a range of empirical analyses. The literature on monetary policy rules has become quite explicit about 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 to such modifications.[9] Within the economic growth literature, analyses such as those by Ross Levine and David Renelt and by Xavier Sala-i-Martin have modified standard growth regression analysis to account for model uncertainty;[10] Gernot Doppelhofer, Ronald Miller, and Sala-i-Martin, as well as Carmen Fernández, Eduardo Ley, and Mark Steel,[11] have explicitly employed the averaging approach to model uncertainty that we endorse.[12] 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, and so forth. It would therefore be a caricature of the empirical literature to suggest that model uncertainty is generally ignored. Relative to these applied approaches, our analysis, we believe, will have some useful suggestions

6. See, for example, Leamer (1978, 1983); Leamer and Leonard (1983).

7. Important examples include Draper (1995), Raftery, Madigan, and Hoeting (1997), and Chipman, George, and McCulloch (2001).

8. Hansen and Sargent (2001a and b, 2002a and b).

9. Levin and Williams (forthcoming) is a recent contribution that is related to the analysis reported here.

10. Levine and Renelt (1992); Sala-i-Martin (1997).

11. Doppelhofer, Miller, and Sala-i-Martin (2000); Fernández, Ley, and Steel (2001b).

12. See Brock and Durlauf (2001).

for 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.

Although our goals are ambitious, we recognize that there are important limits in the extent to which we have achieved them. The decision-theoretic approach is, in an abstract sense, an extremely appealing way to engage in econometric policy evaluation, but significant questions as to how one would implement the approach remain open. We will discuss some ways of making decision theory and model averaging operational, but much substantial work remains to be done. Finally, we do not claim there is "one true path" for empirical work. Debates over the philosophical merits of Bayesian versus frequentist approaches, for example, 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.

The next section of the paper introduces a basic framework for policy evaluation. The discussion is designed to place policy evaluation in a decision-theoretic framework, which we will then exploit throughout the paper. This section is followed by 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. Next we explore some theoretical implications of model uncertainty for policy evaluation. We then discuss 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. The penultimate section provides two applications of our ideas: to monetary policy rules and to 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. The concluding section is followed by computational and data appendixes.

Decision Theory and Uncertainty

Here we describe a basic decision-theoretic approach to policy evaluation. The abstract ideas laid out here constitute the building blocks of modern statistical decision theory.[13] 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. These ideas have been invoked periodically over time as economists have attempted to place empirical research on a more policy-relevant foundation.[14] 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 policies 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 define precisely what constitutes a model; typically a model will incorporate a particular economic theory or theories as well as various functional form specifications. Although 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 the policymaker 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.

13. Good surveys include Berger (1985), French and Rios Insua (2000), and the classic work by Raiffa and Schlaifer (1961); Cox and Hinkley (1974), chapter 11, is a very well written introduction.

14. Chamberlain (2001) is a nice recent example.

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 seeks to minimize the expected loss; 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

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

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

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

As equations 1 and 2 illustrate, policy analysis is thus straightforward once the loss function $l(p, \theta)$ and the probability density $\mu(\theta \mid d,m)$ are specified. However, it is useful to observe that the sorts of calculations associated with equations 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 about the structure of uncertainty and policymaker preferences under which the second moments are the only moments of $\mu(\theta \mid d, m)$ that affect policy assessment. The appropriateness of these assumptions will differ from context to context, and so they should not be adopted 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, in a growth regression in which the proxy is included as an independent variable, the proxy is statistically significant at the 5 percent level. This assessment does not 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 than to focus on test statistics per se. The relevance of this implication differs across empirical literatures; the literature on monetary policy rules is very much focused on the evaluation of such rules with respect to loss functions.[15] In contrast, the economic growth literature is largely 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 typically equates evidence that a policy is relevant for growth with the statistical significance of its associated regression parameter.[16] We will discuss the use of empirical models to evaluate growth policies in more detail in the penultimate section of the paper.

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 θ, or $\hat{\theta}$. Yet most empirical work reports standard errors of estimates rather than measures of the uncertainty concerning underlying parameters. This is a standard objection that Bayesians make of frequentist approaches to econometrics.[17] The import of this criticism will differ across contexts. The reason 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 the associated estimates is of second-order importance in large samples.[18] We will not focus on this issue.

15. See, for example, Taylor (1999a).

16. Barro and Sala-i-Martin (1995).

17. See, for example, Sims (2002).

18. An important exception holds for unit roots in time series, which has led to considerable controversy over how to implement Bayesian methods and, in turn, 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 judge that the sensitivity of results to priors that

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.[19] 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, such as 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 equations 1 and 2; however, there will turn out to be some dimensions along which model uncertainty may warrant a different treatment.

Expected Loss Calculations under Model Uncertainty

To extend our discussion in the previous section 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 he or she knows that the model is true with a probability of 1. Rather, the policymaker will want to compute expected losses conditioning only on the realized data d. Relative to the expected loss calculation described by equation 1, accounting for model uncertainty means that the expected loss for a given policy should be evaluated under the assumption that the model m is an unknown. This requires modifying the policy evaluation equation (equa-

---

occurs in unit roots is a major reason why Bayesian methods have not become more widespread in this context.

19. 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.

tion 1) so that the expected loss associated with each policy accounts for this; the expected loss associated with a policy that conditions only on the data may be calculated as

$$E[l(p,\theta) \mid d] = \int_{\Theta} l(p,\theta)\mu(\theta \mid d)d\theta, \tag{3}$$

where

$$\mu(\theta \mid d) = \sum_{m \in M} \mu(\theta \mid d,m)\mu(m \mid d). \tag{4}$$

The term $\mu(\theta \mid d)$ 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.

Equation 4 illustrates how one can eliminate the 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 in the statistics literature as model averaging.[20]

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 $\mu(\theta \mid d, m)$. As we have argued, in the presence of model uncertainty, the natural object of interest in policy evaluation is $\mu(\theta \mid d)$. Although it is common practice to evaluate the robustness of $\mu(\theta \mid d, m)$ 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 this 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 $\mu(\theta \mid d)$ renders such questions moot, because the information about θ that is contained in each model specification is integrated into its construction.

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

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

$$\mu(m \mid d) = \frac{\mu(d \mid m)\mu(m)}{\mu(d)} \propto \mu(d \mid m)\mu(m), \tag{5}$$

where $\propto$ means "is proportional to." As equation 5 indicates, the calculation of posterior model probabilities depends on two terms. The first term, $\mu(d \mid m)$, 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 a recent analysis by Andrew Levin and John Williams, which is very much in the spirit of model averaging.[21] In their paper, monetary policy rules are evaluated when a forward-looking model, a backward-looking model, and a hybrid, forward- and backward-looking 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 the policy rules is done using their analogue to equation 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 that 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 and the detailed specification of its statistical structure. (Our approach would also account for parameter uncertainty in the calculation of expected losses, but this is an issue distinct from model uncertainty.)

21. Levin and Williams (forthcoming).

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,[22] the mean and variance of δ, once one has accounted for model uncertainty, are

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

and

$$\begin{aligned} \operatorname{var}(\delta \mid d) &= E(\delta^2 \mid d) - \left[E(\delta \mid d)\right]^2 = \\ &\sum_{m \in M} \mu(m \mid d)\left\{\operatorname{var}(\delta \mid d, m) + \left[E(\delta \mid d, m)\right]^2\right\} - \left[E(\delta \mid d)\right]^2 = \\ &\sum_{m \in M} \mu(m \mid d) \operatorname{var}(\delta \mid d, m) + \sum_{m \in M} \mu(m \mid d)\left[E(\delta \mid d, m) - E(\delta \mid d)\right]^2, \end{aligned} \tag{7}$$

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, $\sum_{m \in M} \mu(m \mid d) \operatorname{var}(\delta \mid d, m)$, 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, $\sum_{m \in M} \mu(m \mid d) \left[E(\delta \mid d, m) - E(\delta \mid d)\right]^2$, 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 $E(\delta \mid d, m)$. To see why this second term is interesting, suppose that $\operatorname{var}(\delta \mid d, m)$ is constant across models. Should one conclude that the overall variance is equal to this same value? In general, one

22. Leamer (1978, p. 118).

should not do so. So long as there is any variation in $E(\delta \mid d, m)$ across models, then var($\delta \mid d, m$) < var($\delta \mid d$); that is, 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,[23] this second term explains why one often finds that the predictions of the effect of a policy grossly underestimate the actual uncertainty associated with the effect.

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 equation 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 of uncertainty matter? We would argue that this is an important implication of some of the work associated with the new behavioral economics,[24] 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,[25] which is based on the following experiment. Individuals are asked to state their preferences across four different lotteries. In lottery 1 the individual receives a cash prize if a red ball is drawn from an urn with fifty red and fifty black balls. In lottery 2 the same prize is awarded if a black ball is drawn from the same urn. In lottery 3 the same prize is awarded if a red ball is drawn from a second urn, which also contains a total of 100 red and black balls, but in this urn the proportion of red to black balls is not specified. In lottery 4 the same prize is awarded if a black ball is drawn from the second urn. 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 because 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

23. Draper (1995).
24. See Camerer (1995) for a superb survey of experiments that challenge aspects of expected utility theory.
25. Ellsberg (1961).

uncertainty, in the sense that lotteries 3 and 4 are associated with a range of probabilities for the proportions of red and black balls.

A range of experimental studies have confirmed that individual preferences reflect a distaste for model uncertainty of the type Ellsberg described.[26] This distaste does not appear to be explained by the possibility that the participants do not understand the rules of conditional probability: Paul Slovic and Amos Tversky found that providing participants with written explanations of why preferring lotteries 1 and 2 to lotteries 3 and 4 is inconsistent with expected payoff maximization does not eliminate the paradox.[27] 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).[28] 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 by Itzhak Gilboa and David Schmeidler on ambiguity aversion.[29] This work has proposed a reformulation of individual preferences so that they reflect a dislike of "ambiguity" as well as of 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 this aversion emerges as a feature of behavior not because of the 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 fifty red and fifty black balls and 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 respect to outcomes within a model.

26. Becker and Brownson (1964).
27. Slovic and Tversky (1974).
28. Curley, Yates, and Abrams (1986).
29. Epstein and Wang (1994); Gilboa and Schmeidler (1989).

Although these experiments do not, of course, 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,[30] such preferences may be formalized through the function

$$(8) \qquad (1-e)\int_{\Theta} l(p,\theta)\mu(\theta \mid d)d\theta + e\left[\sup_{m\in M}\int_{\Theta} l(p,\theta)\mu(\theta \mid d,m)d\theta\right].$$

In this expression, e indexes the degree of ambiguity aversion. When $e = 0$, this expression reduces to our earlier expected loss calculation (equation 3). 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.[31]

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 have motivated the new behavioral economics, and so as a positive matter they may be useful in understanding policymaker preferences. Second, we believe that 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, that standard expected loss calculations can be sensibly made. And, of course, as work by Epstein and Wang and by 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.

30. Epstein and Wang (1994).

31. See Hurwicz (1951) for a remarkable early suggestion that preferences like those in equation 8 could bridge different variants of statistical decision theory; these types of preferences are discussed in Manski (forthcoming).

Relation to Other Work

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

EXTREME BOUNDS ANALYSIS. An important research program on model uncertainty originated with Edward Leamer and includes a strategy for rendering the reporting of empirical results more credible.[32] 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 that stipulates that the relationship between p and y should be regarded as fragile if the sign of $\hat{\delta}_m$ changes across specifications.

Following work by Brock and Durlauf,[33] 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 $-\hat{\delta}_m(\bar{\bar{p}} - \bar{p})$. Leamer's rule means that one will choose to implement the policy if and only if $\inf_{m \in M} \hat{\delta}_m(\bar{\bar{p}} - \bar{p}) > 0$. This illustrates how in two respects Leamer's rule presupposes rather special preferences on the part of the analyst. First, the rule requires that $\hat{\delta}_m$ be 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.

32. Leamer (1983); Leamer and Leonard (1983).
33. Brock and Durlauf (2001).

Extreme bounds analysis has been subjected to serious criticism by a number of authors.[34] The major criticism, 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 does a model that does not appear to be misspecified seems, intuitively, a weak reason to conclude that evidence concerning δ is fragile. This does not, however, invalidate Leamer's deep idea that one needs to account for the fragility of regression findings across specifications, nor does it mean that extreme bounds analysis cannot be adapted in a way to respond to the objection.

Following an argument by Brock and Durlauf,[35] 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 equation 8, where $e = 1$ and $\int_\Theta l(p, \theta)\mu(\theta \mid d, m)d\theta = -\hat{\delta}_m p$. 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, Hansen and Sargent have employed robust decision theory to account for the fact that a policymaker typically does not know the true model of the economy.[36] This work has stimulated a growing literature.[37] 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

34. McAleer, Pagan, and Volker (1983).
35. Brock and Durlauf (2001).
36. Hansen and Sargent (2001a and b, 2002a and b).
37. Examples include Giannoni (2002) and Onatski and Stock (2002).

true model is known only to lie within some local neighborhood of models that surround it. 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, that is, $\mu(m)$. 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.[38] To put it another way, Hansen and Sargent assume that Nature draws a model from the neighborhood set of models in such a way as to maximize the loss to the policymaker. They then set their policy rule so as to minimize that loss 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 on a core model) so as to maximize losses to the policymaker, and the policymaker then 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 with 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 robust control theory tools to bear) 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:

> 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 of our model for perturbations from a source that served macroeconomists well before. . . .[39]

38. Wald (1950).
39. Hansen and Sargent (2001a, p. 523).

Our analysis, in contrast, 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 or 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.[40] As such, it attempts to address the sorts of challenges John Taylor has identified in modern research on monetary policy:

> . . . researchers are using many different types of models for evaluating monetary policy 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.[41]

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 long as 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 those local forms of model uncertainty explored in the new robust control approach to macroeconomics. In fact, the minimax approach has proved extremely important in the development of robust approaches to policy

40. 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.

41. Taylor (1999a, p. 658).

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.

Theoretical Implications

Here 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.[42]

Local Robustness Analysis

We first describe an approach to conducting local robustness exercises in policy design. Recall that the previous general discussion placed primary focus on the role of the posterior density of θ—$\mu(\theta \mid d, m)$ if the model is known, $\mu(\theta \mid d)$ if the model is unknown—in allowing a policymaker to evaluate policies. We initially assume that m is known and ask how perturbations around this initial model affect optimal policy choice. Specifically, we 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 $J[p^*(\alpha), \alpha \mid m]$ denote the value of equation 2 evaluated at this optimal policy choice. For technical simplicity we assume that both $J[p^*(\alpha), \alpha \mid m]$ and $p^*(\alpha)$ are twice differentiable.

42. The specific approach we take is further developed in Brock and Durlauf (forthcoming).

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 the way that is least favorable to the policymaker. Metaphorically, one can suppose that the policymaker faces an "adversarial agent," who chooses the actual parameter from this interval in such a way as 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 the one that is optimal when confronted by 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

$$\begin{aligned} & J[p^*(\bar{\alpha}+\Delta), \bar{\alpha}+\Delta \mid m] \\ & \approx J[p^*(\bar{\alpha}), \bar{\alpha} \mid m] \\ & + \left\{ \frac{\partial J[p^*(\bar{\alpha}), \bar{\alpha} \mid m]}{\partial p} \frac{\partial p^*(\bar{\alpha})}{\partial \alpha} + \frac{\partial J[p^*(\bar{\alpha}), \bar{\alpha} \mid m]}{\partial \alpha} \right\} \Delta \\ & \approx J[p^*(\bar{\alpha}), \bar{\alpha} \mid m] + \frac{\partial J[p^*(\bar{\alpha}), \bar{\alpha} \mid m]}{\partial \alpha} \Delta. \end{aligned} \tag{10}$$

The second equality follows from the envelope theorem. Hence the adversarial agent will, for small Δ, choose $\bar{\alpha} + \Delta$ if $\{\partial J[p^*(\bar{\alpha}), \bar{\alpha} \mid m]\}/\partial \alpha > 0$, and $\bar{\alpha} - \Delta$ otherwise.

The robust policy response can thus be computed as a response to the action of the adversarial agent. It is straightforward to show that the robust policy response to the introduction of the adversarial agent is[43]

43. A more detailed analysis can be found in Brock and Durlauf (forthcoming).

$$
(11)\quad
\begin{aligned}
&\mathrm{d}p^* = -\left\{\frac{\partial^2 J[p^*(\bar{\alpha}),\bar{\alpha}\mid m]}{\partial p \partial \alpha}\bigg/\frac{\partial^2 J[p^*(\bar{\alpha}),\bar{\alpha}\mid m]}{\partial p^2}\right\}\Delta \\
&\qquad \text{if } \frac{\partial J[p^*(\bar{\alpha}),\bar{\alpha}\mid m]}{\partial \alpha} > 0 \\
&\mathrm{d}p^* = \left\{\frac{\partial^2 J[p^*(\bar{\alpha}),\bar{\alpha}\mid m]}{\partial p \partial \alpha}\bigg/\frac{\partial^2 J[p^*(\bar{\alpha}),\bar{\alpha}\mid m]}{\partial p^2}\right\}\Delta \\
&\qquad \text{if } \frac{\partial J[p^*(\bar{\alpha}),\bar{\alpha}\mid m]}{\partial \alpha} < 0.
\end{aligned}
$$

One important feature of these formulas is that they indicate how introducing an adversarial agent and considering robustness are 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 random variable with mean zero 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 mean-zero 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 in the special case where $\{\partial J[p^*(\bar{\alpha}), \bar{\alpha} \mid m]\}/\partial\alpha = 0$. The reason is quite intuitive: the presence of the adversarial agent ensures that 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 the optimal choice of policies in the presence of uncertainty.[44] 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

44. Brainard (1967).

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

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

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

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 options 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 trade-off.

How does a robustness analysis affect these calculations? To explore 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 equation 13, it is straightforward to verify that the adversarial agent will choose $\bar{\sigma}_{11} + \Delta$, since the policymaker's payoff is decreasing in the variance of the policy instrument's parameter; that is, 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

$$\frac{dp_1^*(\bar{\sigma}_{11})}{d\sigma_{11}} = -\frac{p_1^*(1+\bar{\sigma}_{22})}{(1+\bar{\sigma}_{11})(1+\bar{\sigma}_{22}) - (1+\bar{\sigma}_{12})^2}, \tag{16}$$

and

$$\frac{dp_2^*(\bar{\sigma}_{11})}{d\sigma_{11}} = -\frac{dp_1^*(\bar{\sigma}_{11})}{d\sigma_{11}}\frac{1+\bar{\sigma}_{12}}{1+\bar{\sigma}_{22}}. \tag{17}$$

Equations 16 and 17 illustrate several basic ideas. First, policy p_2 is always adjusted in the opposite direction from p_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; that is, 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 $|p_1^*|$. This makes intuitive sense: the less trustworthy control is used less aggressively. Combined with $1 + \bar{\sigma}_{12} > 0$, one derives a "precautionary principle" for policymakers: one robustifies against uncertainty in p_1 by using that policy less aggressively and p_2 more aggressively.

Third, this discussion illustrates the difference between evaluating the introduction of risk and evaluating 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. Analogous results may be developed in the context of monetary policy rules. This can be seen using a model by Lars Svensson,[45] which is a one-equation version of an important output-inflation model due to Glenn Rudebusch and Svensson.[46] 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 independently and identically distributed (i.i.d.) sequence of shocks. The inflation gap evolves according to

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

45. Svensson (1996).
46. Rudebusch and Svensson (1999).

where $\phi = 1$. This equation is a proxy for the actual policy process: the policymaker is assumed to be able to control the output gap. The policy-maker'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 $k = \frac{1}{2}\left[1 - \frac{\lambda(1-\beta)}{\beta\delta^2} + \sqrt{1 + \frac{\lambda(1-\beta)}{\beta\delta^2} + \frac{4\lambda}{\delta^2}}\right]$.

To see how robustness works for this model, consider the coefficient ϕ, in equation 18, which Svensson assumes equals 1 in all periods. 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 that the least favorable possible coefficient in N is $1 + \Delta$. (Intuitively, a policymaker prefers less persistence in the inflation process, because it diminishes the net costs to an expansionary policy today.) The optimal choice of the output gap will then equal

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

which is more aggressive than the original rule. To understand the difference, consider that 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 value $\phi = 1 + \Delta$. 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.[47]

Alternatively, robustness may be sought with respect to the measure of control strength δ; that is, rather than treat the control strength as a fixed $\bar{\delta}$, an adversarial agent chooses the measure of control strength from the neighborhood $[\bar{\delta} - \Delta, \bar{\delta} + \Delta]$. One can show that the least favorable parameter for the policymaker in this neighborhood is $\bar{\delta} - \Delta$. This is unsurprising because a smaller value for δ in equation 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, the policymaker will be more aggressive than 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 equation 20. On the other hand, if this term is negative, the coefficient will be smaller than appears in equation 20. Why does 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 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 Taylor:

> I think it is clear that the Phillips curve and the low estimate of the natural rate of unemployment helped lead [in the 1960s] 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.[48]

Suppose one interpreted Taylor as saying that policymakers in the late 1960s and early 1970s had high confidence in their Phillips curve slope estimates—Δ was close to zero. As confidence waned and Δ became larger during the stagflation experience of the 1970s, 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 as Federal Reserve chairman such as Paul Volcker or Alan Greenspan.

47. See Bernhard (2002) for an extended analysis that formalizes this intuition, and Giannoni (2002) for a range of interesting findings along these lines.

48. Taylor (1996, p. 185).

Of course, we do not claim that such a simple model can explain U.S. monetary policy history over the last twenty-five years. We offer this scenario only 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.

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 $\mu(\theta \mid d)$ may still be done using the formula in equation 11 if $J[p^*(\alpha),\alpha \mid m]$ is replaced with $J[p^*(\alpha),\alpha]$, where

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

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

We will use equation 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 equation 8, although one certainly can conduct the analysis under such preferences.

APPLICATION TO GROWTH ECONOMICS. To see what new insights emerge when one introduces multiple core models, we develop an analysis of robustness in a growth context. We first 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 the rate of economic growth in country i. 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 for country i be

$$g_i = v_m' S_{i,m} + \delta_m p_i + \varepsilon_i = \delta_m p_i + \upsilon_{im}. \tag{23}$$

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

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

where *r* scales the relative weights of expected growth and the uncertainty in growth (as measured by the standard deviation). The optimal policy level for a given model will, under these preferences, equal

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

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, and 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 $[\bar{\sigma}_{\delta\delta} - \Delta, \bar{\sigma}_{\delta\delta} + \Delta]$. That 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

$$(26) \qquad \mathrm{d}p^* = -\frac{p^*}{\sigma_{\delta\delta}}\Delta,$$

which means that the robust policy level $p^* + \mathrm{d}p^*$ is smaller than $p^*(\sigma_{\delta\delta})$ if $p^* > 0$, and $p^* + \mathrm{d}p^*$ 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 expression 24, one can show from equation 25 that p^* is increasing in $E(\delta)$, that $|p^*|$ is decreasing in $\sigma_{\delta\delta}$, that p^* is decreasing in $\sigma_{\delta\upsilon}$, and that p^* is decreasing in *r* if $E(\delta) > 0$ but increasing in *r* if $E(\delta) < 0$.

Relative to these results, in particular equations 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 equations 6 and 7. Once one has replaced the model-dependent moments in equation 25 with the moments described by equations 6 and 7, one can 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 equation 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

$$\text{(27)} \qquad \frac{\mathrm{d}J}{\mathrm{d}\sigma_{\delta\delta,1}} = -\frac{r}{2}\mu(m=1)p^{*2}\,.$$

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

$$\text{(28)} \qquad \mathrm{d}p^* = \frac{-\mu(m=1)p^*}{\bar{\sigma}_{\delta\delta}}\Delta.$$

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 $\mu(m = i)$ 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

$$\text{(29)} \qquad \frac{\mathrm{d}J}{\mathrm{d}\mu(m=1)} = J_1 - J_2.$$

This formula indicates that if one is considering robustness with respect to posterior model probabilities in the interval $[\mu(m = 1) - \Delta, \mu(m = 1) + \Delta]$, 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 the policymaker prefers model 1, conditional on p^*. In this case the optimal policy response will follow

$$\text{(30)} \qquad dp^* = \frac{1}{r\sigma_{\delta\delta}}\left[E_1(\delta) - E_2(\delta) - rp^*(\sigma_{\delta\delta,1} - \sigma_{\delta\delta,2}) - r(\sigma_{\delta\upsilon_i,1} - \sigma_{\delta\upsilon_i,2})\right]\Delta,$$

where E_m denotes the expectation under model m.

In general, it is unclear whether the change described by equation 30 is positive or negative. However, if r is small, then $dp^* \approx -[E_1(\delta) - E_2(\delta)]\Delta$, so that if $E_1(\delta) > E_2(\delta)$, then the policy is used less aggressively when robustness is incorporated into the policy construction.

Issues in Empirical Implementation

Here 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; the next section will present some empirical exercises.

Bayesian, Frequentist, and Waldean Approaches to Model Evaluation

From the perspective of empirical analysis, the key objects that must be computed are $\mu(\theta \mid d, m)$ and $\mu(m \mid d)$. These calculations require that a researcher take a stance on the use of Bayesian versus frequentist methods. Here 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 has specified a way of estimating $\mu(\theta \mid d, m)$ that formally accounts for prior information. To see this, notice that

$$\text{(31)} \qquad \mu(\theta \mid d, m) = \frac{\mu(\theta, d \mid m)}{\mu(d \mid m)} = \frac{\mu(d \mid \theta, m)\mu(\theta \mid m)}{\mu(d \mid m)},$$

or

$$\text{(32)} \qquad \mu(\theta \mid d, m) \propto \mu(d \mid \theta, m)\mu(\theta \mid m).$$

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: $\mu(d \mid \theta, m)$, the probability of the data *d* given θ, and μ(θ | *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 of course is natural when the unknowns are shocks). Rather, the uncertainty about θ is subjective, in that it is characterized relative to the policymaker.

This formulation is what David Lindley has called "the complete Bayesian paradigm," concluding as follows:

> 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. . . .[49]

Lindley's distinction between evaluating probabilities and calculating them 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; some 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,

> 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 dishonest 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. . . .[50]

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

49. Lindley (1990, p. 54).
50. Leeper, Sims, and Zha (1996, p. 5).

MODEL UNCERTAINTY AND FREQUENTIST METHODS. Although 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 $l(p \mid d,m)$. 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, that is, by computing an expected loss of the form

$$E\left[l(p \mid d)\right] = \sum_{m \in M} l(p \mid d,m)\mu(m \mid d) \propto \sum_{m \in M} l(p \mid d,m)\mu(d \mid m)\mu(m). \tag{33}$$

Although 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, $\mu(d \mid m)$ may be approximated by the standard likelihood statistic.[51]

Although an orthodox Bayesian might object to analyses such as equation 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.[52] In our empirical applications we will use both full Bayesian and pseudo-Bayesian strategies to illustrate how each may be made operational.

THE WALDEAN APPROACH. Perhaps the major non-Bayesian approach to decision theory is due to Abraham Wald. In this type of analysis, the

51. We say "approximated" because 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.

52. Similarly, strict Bayesians object to empirical Bayes methods (Maritz and Lwin, 1989), but such methods have proved useful in practice, despite alleged philosophical shortcomings; see Rubin (1980).

focus is on the development of statistical decision functions, that is, 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 R associated with a given statistical decision function:[53]

$$R(p,\theta) = \int_D l[p(d),\theta]\mu(d \mid \theta)\mathrm{d}d. \tag{34}$$

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:

$$E[R(p,\theta)] = \int_\Theta \left\{ \int_D l[p(d),\theta]\mu(d \mid \theta)\mathrm{d}d \right\} \mu(\theta)\mathrm{d}\theta. \tag{35}$$

By a standard calculation,[54] the evaluation of average risk leads to the same expected loss calculation as in equation 1 when one uses the complete Bayesian solution we have described; $\mu(\theta)$ functions as a prior density.

A meaningful contrast between the Waldean and Bayesian approaches occurs if instead one follows a minimax strategy: choose $p(d)$ so as to minimize

$$\max_{\theta\in\Theta} \int_D l[p(d),\theta]\mu(d \mid \theta)\mathrm{d}d. \tag{36}$$

Are there cases where the Waldean 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 Waldean approach with respect to cross-model uncertainty regardless of how one evaluates within-model uncertainty.

53. Cox and Hinkley (1974, p. 429).
54. Raiffa and Schlaifer (1961, p. 15).

Characterizing Model Uncertainty

SPECIFYING ELEMENTS OF THE MODEL SPACE. The specification of a space of possible models is ultimately a matter of the 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 about models in economics 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, and so forth are a good illustration of the limits to which the current state of economic theory can guide a policymaker. Of course, the persistence of disagreements over fundamental aspects of the economy also reflects the absence of empirical evidence that would be decisive in adjudicating among alternative theories. At the same time, there are in most policy-relevant cases a rich range of alternative theories whose empirical analogues 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 things subject to specification uncertainty in macroeconomic contexts include the 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, many empirical proxies 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 data generated 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." For example, one can interpret heterogeneity uncertainty in many cases as a question of incorporating nonlinearity, and thus as a form of specification uncertainty. Our purpose is merely to indicate some of the judgments that need to be made in constructing a model space.

INTERPRETING A MODEL SPACE. Although the specification of a model space is something that may be discussed only 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 distinguish environments that are M-closed and M-open:[55] 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 increases, the "true" model will receive an asymptotic weight of 1;[56] 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.

Although the asymptotics of statistical procedures that account for model uncertainty are reasonably well understood for both 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 that economic policymakers typically face, 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,

55. Bernardo and Smith (1994).

56. This is true so long as appropriate prior coefficient densities are used; see Fernández, Ley, and Steel (2001a) for discussion.

to minimize 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 apply only to models that 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, because it requires thinking about decisions when the decisionmaker does not know the support of the uncertainty he or she faces. Although recent work in economic theory has addressed some aspects of this problem, it is far from well understood.) Further, because 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 are related to the general questions, which continue to be debated in Bayesian contexts, concerning the nature of prior probabilities. 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 an excessive impact on the posterior description of the parameters.[57]

—*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,[58] robustness may be interpreted as a safeguard against misspecification of prior information.

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

57. This is the property defended by Leeper, Sims, and Zha (1996) in their discussion of priors cited earlier.

58. Berger (1993, ch. 4).

Of course, these obviously desirable properties leave a great deal of discretion to a given researcher. And one could easily add other desiderata. The arguments made by Leeper, Sims, and Zha,[59] 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 in part a nuisance, whose choice should be regarded as no 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 because 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 in a case 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 equation 9, and one to which we will return in the context of growth econometrics in the next section. The recent efforts to employ model averaging to account for uncertainty with respect to variable inclusion generally assume that the possible models are all equally likely a priori.[60] 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 a probability of 2^{-K}. This is equivalent to assuming that the probability that a given variable is present in the "true" model is equal to 0.5 and is independent of the presence or absence of any of the other regressors in the model. Some have proposed altering the probability of variable inclusion in order to give greater weight to models with a small number of regressors,[61] 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,[62] but the independence assumption is, at least in our reading, essentially universal.

59. Leeper, Sims, and Zha (1996).

60. For example, Fernández, Ley, and Steel (2001a); Raftery, Madigan, and Hoeting (1997).

61. See Doppelhofer, Miller, and Sala-i-Martin (2000).

62. Brown, Vannucci, and Fearn (1998, 2002).

As argued by Brock and Durlauf,[63] such a formulation of priors on the model space is difficult to justify. For example, the growth theory that holds 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 that 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 issue in discrete choice analysis is how the probability of an individual choosing a red bus over a taxi is affected by adding the possibility of choosing a blue bus. Under the independence of irrelevant alternatives assumption 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 because the blue bus is a far closer substitute for 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.

Empirical Applications

Monetary Policy Rules

Our first empirical 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 Taylor on what are now called 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; we assume that monetary policy is constrained to follow a Taylor rule; and we consider only backward-looking models. We compute estimates of the effects of alternative choices of monetary policy parameters and contrast those estimates with those of the well-known Rudebusch and Svensson model.[64]

63. Brock and Durlauf (2001).
64. Rudebusch and Svensson (1999).

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.[65] The same concern with robustness appears in a number of papers in the Taylor volume and in recent papers such as one by Levin and Williams.[66] 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 that 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 that the weights are simply the model likelihood: well-fitting models get more weight than do ill-fitting models. We view our approach as complementing rather than replacing 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 considering 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 trade-off between, for example, raising interest rates 50 basis points this month, and raising them 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, or $\mu(m)$'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 simultane-

65. McCallum (1988).
66. Taylor (1999b); Levin and Williams (forthcoming).

ity 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 by Onatski and Noah Williams.[67] The research presented here is intended both to complement this work and to illustrate the frequentist approach to model averaging (equation 33) in a simple context.

We employ the same notation as previously: y_t is the output gap; π_t is quarterly inflation, at an annual rate; 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

$$R = \text{var}(\pi_t) + \lambda_y \,\text{var}(y_t) + \lambda_i \,\text{var}(\Delta i_t). \tag{37}$$

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,[68] nor do we allow the weights to vary across specifications.

We consider different three-equation systems for i_t, y_t, and π_t. Our specification assumes that the output gap and the 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 equation 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 the IS curve and the equation with π_t on the left-hand side 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

67. Onatski and Stock (2002); Onatski and Williams (2003).
68. See Woodford (2002).

to sum to unity. The most profligate specification entails four lags of *i, y,* and π in both equations, which is almost but not quite an unrestricted vector autoregression ("almost" because lags of inflation in the Phillips curve are always constrained to sum to 1).

Specifically, we vary lags across specifications as follows. In the IS curve we include specifications of two types. First, we construct 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 (where 1 designates the most recent previous period), 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

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

The first two terms on the right-hand side of equation 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, $i_{t\text{-}1} - \pi_{t\text{-}1}$, was always present, with lags of i adjusted to prevent linear dependence in the regressors in particular versions of equation 39. This also produces $4 \times 5 \times 4$ specifications.

For 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

$$\begin{aligned} \pi_t &= \beta_{\pi 1} \pi_{t-1} + \beta_{y1} y_{t-1} \\ &+ \left[\sum_{j=2}^{4} \beta_{\pi j} \pi_{t-j} + \sum_{j=2}^{4} \beta_{yj} y_{t-j} + \sum_{j=1}^{4} \beta_{ij} i_{t-j} \right] + v_t, \sum_{j=1}^{4} \beta_{\pi j} = 1. \end{aligned} \tag{40}$$

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

Each of the regressions described was estimated alternately with a constant term as the only deterministic component and with a constant term as well as a dummy for periods after 1984:1. The dummy is intended to crudely allow for changes initially documented by Margaret McConnell and Gabriel Perez-Quiros.[69] Thus the total number of specifications is $[(4 \times 5 \times 4) + (4 \times 5 \times 4)] \times 4 \times 4 \times 5 \times 2 = 25{,}600$, 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's sake.[70]

For each model we estimate the IS and Phillips curves by ordinary least squares (OLS). In conjunction with choices of g_π, g_y, and g_i, in equation 38, one can compute estimates of the total loss described by equation 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 Bayesian information criterion–adjusted likelihood, L_m. We compute model average risk as

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

This equation fits into the frequentist approach outlined in the previous section, with $\hat{R}_m$ playing the role of $l(p \mid d, m)$ and $L_m / \sum_{m \in M} L_m$ the role of $\mu(d \mid m)$ in equation 33, under the assumption that all models have equal prior probability, that is, that $\mu(m) = 1/25{,}600$.

To clarify and illustrate the effects of model averaging, we contrast our model averaging results with those of one well-known special case of the class of models considered. This is the Rudebusch and Svensson model.[71] In this model the IS equation is

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

69. Statistical tests in McConnell and Perez-Quiros (2000) suggested that a shift in GDP and some of its components occurred around 1984.

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

71. Rudebusch and Svensson (1999).

and the Phillips curve equation is

$$\pi_t = \beta_{\pi 1}\pi_{t-1} + \beta_{\pi 2}\pi_{t-2} + \beta_{\pi 3}\pi_{t-3} + \beta_{\pi 4}\pi_{t-4} + \beta_{y1} y_{t-1} + v_t, \quad (43)$$

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 in equation 37, we solve for Taylor rule parameters that were optimal under the Rudebusch and Svensson model. We compute risk according to the model, denoting it as $\hat{R}_{RS}$, as well as according to all other models in the model space we have described. The model-specific risk calculations are then averaged according to equation 33 to produce model average risk. The objective of this exercise is to see whether the Rudebusch and Svensson figures for risk well match those for model averages. The ranges of values for the risk parameters are those suggested by Levin and Williams:[72] $\lambda_y = \{0.0, 0.5, 1.0, 2.0\}$ and $\lambda_i = \{0.1, 0.5, 1.0\}$, twelve sets of values in all.

Apart from lags, the sample period is 1969:1–2002:4. Inflation is computed as annualized growth in the GDP deflator, and the output gap is computed from actual real GDP and the Congressional Budget Office's estimate of potential GDP. We use the latest data available, thus abstracting from possible complications from data revision.

Results are presented in table 1. The values given for the Taylor rule parameters g_π, g_y, and g_i are those that are optimal under Rudebusch and Svensson, as found by a grid search. These display a familiar and intuitive pattern: larger weights on output volatility (higher λ_y) lead to higher optimal g_y, and larger 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 table 1, we compute model average risk *ER* based on equation 41 and compare it with 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

72. Levin and Williams (forthcoming).

Table 1. Effects of Model Uncertainty on Risk for Optimal Rudebusch-Svensson Rules

Risk parameters[a]		*Taylor rule parameters*[b]			*Change in risk*[c] *(percent)*
λ_y	λ_i	g_π	g_y	g_i	
0.0	0.1	4.5	2.0	0.2	306
0.0	0.5	2.3	1.0	0.4	17
0.0	1.0	1.7	0.7	0.5	–2
0.5	0.1	4.4	2.7	0.0	56
0.5	0.5	2.4	1.3	0.3	1
0.5	1.0	1.8	0.9	0.4	–9
1.0	0.1	4.3	3.2	–0.1	16
1.0	0.5	2.5	1.6	0.2	–7
1.0	1.0	1.7	1.0	0.4	–10
2.0	0.1	4.1	3.7	–0.2	8
2.0	0.5	2.5	1.9	0.1	–13
2.0	1.0	1.8	1.3	0.3	–14

Source: Authors' calculations.

a. Assumed weights on the variances of y_i (λ_y) and Δi_i (λ_i) in the equation for the model average risk (equation 37 in the text).

b. Optimal values for the parameters π_t (g_π), y_t (g_y), and i_{t-1} (g_i) in the monetary policy rule (equation 38 in the text) when the Rudebusch-Svensson (1999) model given by equations 42 and 43 is assumed to generate the data.

c. Increase in risk when the model average risk (equation 37) is used rather than the risk estimated using the Rudebusch-Svensson model; the increase is calculated as $100 \times [(ER / \hat{R}_{RS}) - 1]$, where ER is defined as in equation 41.

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. Although the figure in the last column of the first row of table 1 is quite large (over 300 percent), the other numbers are much smaller and scattered fairly evenly around zero.

We take this as illustrating two points. First, when our results are compared with those of Levin and Williams,[73] it seems that there is substantially less variation in risk within the class of backward-looking models we have studied than there is between backward- and forward-looking models. Specifically, findings for the Rudebusch-Svensson baseline are generally representative of the risk associated with the monetary policies considered in the table. The second and potentially more useful point from the perspective of future research is one emphasized in our discussion above: model averaging allows tractable accounting for the effects of model uncertainty.

73. Levin and Williams (forthcoming).

Economic Growth

In our second application, which will follow the full Bayesian approach we discussed in the previous main section, we turn to the empirical growth literature. Our analysis focuses 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.[74]

GROWTH ECONOMETRICS: GENERAL ISSUES. Much 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.[75] From the vantage point of using such regressions to evaluate a growth policy p, a canonical form of this regression is

$$g_i = \beta' X_i + \gamma' Z_i + \delta p_i + \varepsilon_i, \tag{44}$$

where g_i is real growth per capita across some fixed time interval, X_i is a set of regressors suggested by the Solow growth model (initial population growth, technological change, and physical and human capital saving 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. Edmond Malinvaud summarized the importance of such regressions in policy analyses:

> 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. . . .[76]

Regressions such as equation 44 have been used to evaluate many different policies: a survey of this type of empirical work can be found in Barro and Sala-i-Martin.[77] For our purposes the main point is that the evaluation of a growth policy typically amounts to assessing the statistical

74. See Brock and Durlauf (2001), Durlauf (2000), and Temple (2000) for related analyses.
75. See Barro (1991) and Mankiw, Romer, and Weil (1992).
76. Malinvaud (1998, p. 781).
77. Barro and Sala-i-Martin (1995).

significance of δ for a baseline specification like that in equation 44 and a small set of alternative specifications, which typically amount to changing the variables included in Z_i. Such analyses pay only indirect and unsystematic 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 elsewhere.[78] 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 found nearly as many alternative growth theories and associated empirical measures as there are countries in the standard data sets;[79] by now the number of theories exceeds the number of countries. This plethora of alternative theories is particularly worrisome because, following Brock and Durlauf,[80] growth economics suffers from theory open-endedness. Theory open-endedness 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 those of equation 44 are interpretable for policy evaluation only to the extent that the regression specification is considered sufficiently rich that the data from each country constitute a draw from the common statistical model defined by the regression. Although 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 an adviser to a sub-Saharan African government on some policy wishes to use 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 are the same for countries in sub-Saharan Africa as for the United

78. 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.

79. Durlauf and Quah (1999).

80. Brock and Durlauf (2001).

States? 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 equation 44 to uncover growth determinants.[81]

A number of studies have documented parameter uncertainty of various forms.[82] 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.[83]

A third problem is that it is far from clear that statistical significance can provide a useful guide to policy evaluation. Although the abstract argument was made earlier in this paper, it is particularly salient in the growth context, and so we expand upon it here. Suppose one's 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,[84] one can think about t statistic rules from a decision theory perspective. A simple way to do this is to interpret a t statistic rule as implying that, when comparing $\bar{p}$ with 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

81. 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 property 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.

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

83. 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 income per capita, for example, with some growth determinant. It seems fair to say that these efforts are generally ad hoc.

84. Brock and Durlauf (2001).

to its standard deviation, one can approximate the t statistic rule as implying that one makes the policy change based upon

$$\text{(45)}\quad \begin{aligned} &E\left[l\left(\overline{\overline{p}},\theta\right)\middle| d,m\right]-E\left[l\left(\overline{p},\theta\right)\middle| d,m\right]\\ &=\left[-E\left(\delta\middle| d,m\right)+2\operatorname{var}\left(\delta\middle| d,m\right)^{1/2}\right]\left(\overline{\overline{p}}-\overline{p}\right), \end{aligned}$$

with the policy change adopted only if the value of equation 45 is less than 0. (If we were considering instead a reduction in the policy variable, the requirement would be that the value of equation 45 be 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 trade-off between the mean and the variance of the policy effect.[85]

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 equation 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. To show how one might address these problems, we consider a particular policy question: should the countries of sub-Saharan Africa lower tariffs in order to improve their growth performance?[86] Our analysis based only on cross-country growth regressions will obviously be a caricature of the actual policy process, as it ignores the plethora of information available to organizations such as the World Bank to help inform policy decisions, but

85. Blinder (1997, p. 6) makes a similar criticism of the use 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, 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.

86. Data are available for thirty-one countries in sub-Saharan Africa across the time period we consider: Benin, Botswana, Burkina Faso, Burundi, Cameroon, Central African Republic, Congo, Côte d'Ivoire, Ethiopia, Gabon, Ghana, Kenya, Madagascar, Malawi, Mali, Mauritania, Mauritius, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, Somalia, South Africa, Sudan, Tanzania, Togo, Uganda, Zaire, Zambia, and Zimbabwe.

for expositional purposes we treat such an analysis as if it were the sole basis on which policy decisions are made.

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 African 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 equation 44, this means that every element in the model space contains X_i and p_i.[87] We then introduce six possible additional categories of theories of cross-country growth differences that have received prominence in the literature: exchange rate policies, government spending policies, inflation, characteristics of the economic system, characteristics of the financial system, and characteristics of the political system. The first three categories, roughly speaking, represent theories that relate various government policies to economic growth. The second three represent theories that link growth to longer-run structural aspects of a country's economic and political system. Although these categories do not exhaust the range of new growth theorizing, we argue that they cover a relatively broad part of the spectrum.

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 such 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

87. 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 Fernández, Ley, and Steel (2001b), where the focus is on identifying which variables should be included in a growth model out of 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 in that we include the Solow variables and the tariff variable with a probability of 1.

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 it is no different from the judgments often necessary to implement models such as the nested logit.[88]

Second, once one has specified a set of theories, it is necessary to specify how the various theories are to be 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 appendix B. For each of these sets of variables, we allow each nonempty 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 which models with parameter heterogeneity take the form

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

where $\xi_{i,\mathrm{SSA}}$ is an indicator variable that equals 1 if country i is in sub-Saharan Africa and 0 otherwise. This type of heterogeneity has proved useful in previous work on sub-Saharan Africa: for example, Brock and Durlauf 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.[89]

88. George (1999) suggests 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.

89. Brock and Durlauf (2001); Easterly and Levine (1997).

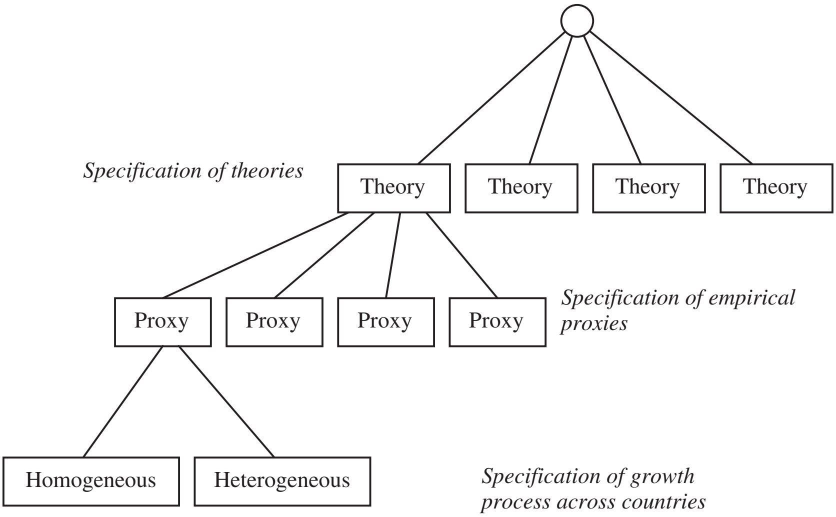
Figure 1 illustrates our formulation of model uncertainty for growth regressions. The first level of uncertainty that must be resolved in defining a particular model concerns the set of growth theories to include in the specification. The second level of uncertainty that must be resolved is which empirical proxies for these theories are to be used. Once a set of theories and associated empirical proxies are specified, the final level of uncertainty that must be resolved is whether or not sub-Saharan Africa obeys a different growth process from the rest of the world. If one were to enumerate every ramification of the tree diagram in 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 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 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 also included in that model. Second, we assign equal probability weights to each of the possible empirical analogues of a theory (that is, to each combination of variables used to instantiate 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 equation 46; we allow the error variances for sub-Saharan African countries to differ from those of the rest of the world. For each pair of corresponding models with and without heterogeneity, we assign a probability of q to the heterogeneous model and of $1-q$ to the homogeneous model. For expositional purposes, we report $q = 0$ separately. Overall, there are 4,096 different models generated by theory uncertainty and regressor choice uncertainty; allowing for heterogeneity uncertainty doubles this to 8,192.

This tree structure for the probabilities represents an effort to address a problem in previous work,[90] namely, 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 probability of inclusion. Our

90. See Brock and Durlauf (2001), Doppelhofer, Miller, and Sala-i-Martin (2000), and Fernández, Ley, and Steel (2001b).

Figure 1. Construction of the Model Space for Analyzing Growth



approach is designed to distinguish the questions of uncertainty over theories from questions of uncertainty concerning empirical proxies. Although our approach is, we believe, an improvement on previous ways of assigning prior probabilities, we fully expect that it will evolve in future work.[91]

To compute posterior densities for the parameters and associated expected growth levels in the models defined by equations 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 appendix A, 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 pri-

91. 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 such a way that the empirical proxies are combined to form an optimal estimate of the empirical "signal" associated with the theory.

marily for expositional clarity; see Fernández, Ley, and Steel for an extensive discussion of the appropriate use of priors in linear model averaging contexts.[92]

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 and capital goods and corresponds to the variable OWTI (see appendix B) in the standard Barro and Lee data set. The first data column reports the OLS estimate of the coefficient on the tariff variable based on regressions that include only the Solow variables and the tariff variable. The second column reports the OLS estimate when all available variables are included. The next three columns report Bayesian model averaging exercises under different theory inclusion probabilities q; we consider $q = 0.25$, 0.5, and 0.75. The next two columns report those estimates among all models estimated in the Bayesian model averaging analysis that produce the minimum and maximum posterior means of the parameter. The final two columns report the results for the analogous models whose payoffs under the t statistic rule in equation 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 coefficients 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, in which all prior model probability is assigned to one model.) The last four columns 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 q produce very little difference in posterior means and standard deviations. The model averaging estimates of the mean of the tariff parameter are more than 10 percent higher than the tariff coefficient of the OLS regression for the narrow Solow model and about 3 percent larger than that for the expanded Solow model. The standard deviations from our model averaging estimates are about 10 percent smaller than that for 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

92. Fernández, Ley, and Steel (2001a).

Table 2. Tariff Parameter Moments under Alternative Theory Inclusion Probabilities[a]

			Results of model averaging exercise						
			Model averaging estimates under indicated inclusion probability (*q*)[d]						
Statistic	*OLS*[b]	*Full OLS*[c]	*0.25*	*0.50*	*0.75*	*Model estimate with minimum coefficient*[e]	*Model estimate with maximum coefficient*[f]	*Model estimate with minimum coefficient + 2 SE*[g]	*Model estimate with maximum coefficient + 2 SE*[h]
Mean	–0.5377	–0.5725	–0.6022	–0.5992	–0.5946	–0.6508	–0.3595	–0.6332	–0.3737
Standard deviation	(0.2282)	(0.1977)	(0.1920)	(0.1885)	(0.1870)	(0.2163)	(0.2222)	(0.1898)	(0.2297)

Source: Authors' calculations.

a. The tariff variable measures tariffs on intermediate and capital goods and corresponds to variable OWTI in the Barro and Lee (1994) data set. The reported coefficient indicates the change in the tariff that results in optimal growth. The dependent variable is annual growth in GDP.

b. Ordinary least-squares estimates of the coefficient on the tariff variable based on equation 44 in the text, where only the tariff variable and the Solow variables are included in the regression.

c. Ordinary least-squares estimates in which all available policy and structural variables are included in the regression in addition to the tariff and Solow variables.

d. Bayesian model averaging estimates over alternative specifications of equation 44, in which priors are generated with the alternative inclusion probabilities indicated.

e. Estimate for that specification of equation 44 with the smallest coefficient estimate among all models used in the model averaging exercise; includes a constant term and the variables MNGD, MINV, MSCH, MGDP60, OWTI, RERD, GGCFD, CIVILLY, DMCYBL (see appendix B for variable definitions).

f. Estimates for that specification of equation 44 with the largest coefficient estimate of the models used in the Bayesian model averaging exercise; includes a constant term and the variables MNGD, MINV, MSCH, MGDP60, OWTI, BMPL, LLY, EcOrg, RULELAW, CIVILLY.

g. Estimates for that specification of equation 44 where the coefficient estimate plus two standard errors (SE) is the smallest of the models used in the Bayesian model averaging exercise; includes a constant term and the variables MNGD, MINV, MSCH, MGDP60, OWTI, RERD, GVXDXE5, DMCYBL, PI6089, PIHYP6089.

h. Estimates for that specification of equation 44 where the coefficient estimate plus two standard errors is the largest of the models used in the Bayesian model averaging exercise; includes a constant term and the variables MNGD, MINV, MSCH, MGDP60, OWTI, DCPY, LLY, EcOrg, RULELAW, CIVILLY.

deviation less than half the size of the coefficient; by the "*t* statistic" loss function described by equation 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 rather than the OLS estimates. 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 with 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, the column labeled "Minimum coefficient" reports the model that would provide the strongest support for a tariff reduction (that is, has the smallest parameter estimate). A policymaker with *t* statistic preferences would find the model described in the column labeled "Minimum coefficient + 2 SE" 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 cared only about the mean of the parameter conditional on a model would make a policy evaluation on the basis of the model described in the column labeled "Maximum 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 the column labeled "Maximum coefficient + 2 SE."

These results indicate that the policy recommendation implied by the OLS and model averaging exercises is similar to that 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 tariff reduction is strong. 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 or she 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 equation 44 for the tariff parameter from regressions based on the Solow variables plus tariffs (first data column), and regressions based on the Solow variables, tariffs, and all other variables (second column). For the model averaging analysis, we focus on the case where the theory inclusion probability is 0.5, and we consider prior probability weights on models with heterogeneity and corresponding models without heterogeneity to equal 0.5, 0.75, and 1, respectively. These results appear in the last three columns.

The latter results indicate a significant discontinuity in the mean and standard deviation of the tariff parameter for $q = 1$ compared with the other cases. In particular, the first two moments of the parameter are similar to those found in table 2 for $q = 0.5$ and $q = 0.75$; allowing for heterogeneity slightly lowers the posterior mean and raises the posterior standard deviation by about 20 percent for a prior heterogeneity probability of 0.5 and by about 50 percent for a prior heterogeneity probability of 0.75. 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 four times as large as those 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 sub-Saharan African heterogeneity are very small. When $q = 0.5$, the total posterior probability on models with heterogeneity is only 0.014 (not shown); for $q = 0.75$, the posterior probability is only 0.04. 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 very different estimates, which would lead a policymaker with preferences like those expressed in equation 45 to decide against a tariff reduction. Our other regression exercises also lead to a rejection of the tariff reduction under those preferences. In both of the Solow cases, if sub-Saharan African heterogeneity is included with a probability of 1, the standard deviation of the posterior density of the tariff coefficient for sub-Saharan African 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.

Table 3. Tariff Parameter Moments under Alternative Prior Heterogeneity Probabilities

			Prior heterogeneity probability[c]		
Sample	*OLS*[a]	*Full OLS*[b]	*0.50*	*0.75*	*1.00*
Sub-Saharan Africa	–0.4320	–0.2512	–0.6079	–0.6246	–1.2322
	(0.8943)	(1.0112)	(0.2205)	(0.2707)	(0.7678)
Rest of world	–0.6276	–0.4630	–0.5981	–0.5961	–0.5222
	(0.2067)	(0.2005)	(0.1890)	(0.1899)	(0.2067)

Source: Authors' calculations.

a. OLS estimates of the coefficient on the tariff variable based on equation 46 in the text, where only the tariff variable and the Solow variables are available for inclusion in the regression. Standard deviations are reported in parentheses.

b. OLS estimates where all available policy and structural variables as well as the tariff variable and the Solow variables are included in the regression.

c. Bayesian model averaging estimates using versions of equation 46 with priors generated with an inclusion probability of 0.5 and prior probabilities of coefficient heterogeneity as indicated.

We are surprised by the weakness of the evidence on heterogeneity given previous work that found parameter heterogeneity,[93] 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 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 analogue) 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 are compared with and without a 10 percent reduction of tariffs beyond 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 percent 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 is subject to Lucas critique–type arguments, in that we do not account for the effects of the policy change on model parameters (or for

93. Brock and Durlauf (2001).

Table 4. Changes in Moments of the Growth Rate in Sub-Saharan African Countries for a 10 Percent Decrease in Tariffs[a]

		Expected growth			*Growth expected from model with smallest coefficient on tariff variable*			*Growth expected from model with largest coefficient on tariff variable*		
Country	*True value*	*Before experiment*	*After experiment*	*Change*	*Before experiment*	*After experiment*	*Change*	*Before experiment*	*After experiment*	*Change*
Benin	–0.0412	0.1320 (0.3438)	0.1478 (0.3439)	0.0158 (0.00002)	0.2681 (0.3461)	0.2853 (0.3459)	0.0171 (–0.0002)	0.2728 (0.3548)	0.2823 (0.3547)	0.0094 (0.0001)
Botswana	1.3423	1.0215 (0.3437)	1.0344 (0.3438)	0.0129 (0.00006)	1.0361 (0.3592)	1.0501 (0.3595)	0.0140 (0.0003)	1.0268 (0.3586)	1.0345 (0.3588)	0.0077 (0.0002)
Burkina Faso	0.4824	0.5338 (0.3556)	0.5627 (0.3546)	0.0288 (–0.0010)	0.3855 (0.3699)	0.4168 (0.3685)	0.0313 (–0.0014)	0.1983 (0.3689)	0.2157 (0.3675)	0.0173 (–0.0014)
Burundi	–0.1299	–0.0898 (0.3486)	–0.0766 (0.3486)	0.0132 (–0.00003)	–0.3073 (0.3596)	–0.2929 (0.3596)	0.0143 (0.00005)	–0.1940 (0.3608)	–0.1861 (0.3608)	0.0079 (–0.00005)
Cameroon	0.9016	0.5033 (0.3399)	0.5189 (0.3398)	0.0156 (–0.0001)	0.3448 (0.3462)	0.3617 (0.3461)	0.0169 (–0.0001)	0.7529 (0.3529)	0.7623 (0.3527)	0.0093 (–0.0001)
Central Afr. Rep.	–0.0603	0.1417 (0.3381)	0.1536 (0.3381)	0.0119 (0.00007)	0.2866 (0.3458)	0.2996 (0.3458)	0.0130 (0.00001)	0.4198 (0.3485)	0.4270 (0.3486)	0.0071 (–0.00005)
Congo	0.9557	0.6035 (0.3524)	0.6154 (0.3525)	0.0118 (0.0001)	0.7964 (0.3559)	0.8092 (0.3559)	0.0128 (–0.00005)	0.9217 (0.3578)	0.9289 (0.3580)	0.0071 (0.0002)
Côte d'Ivoire	0.2066	0.3445 (0.3457)	0.3574 (0.3456)	0.0129 (–0.00002)	0.0845 (0.3485)	0.0985 (0.3486)	0.0140 (0.0001)	0.5129 (0.3547)	0.5207 (0.3547)	0.0077 (–0.00001)
Ethiopia	0.1317	0.2459 (0.3452)	0.2578 (0.3455)	0.0119 (0.0002)	0.2210 (0.3528)	0.2340 (0.3531)	0.0130 (0.0003)	0.1110 (0.3602)	0.1182 (0.3603)	0.0071 (0.0001)
Gabon	1.4094	0.8484 (0.3400)	0.8613 (0.3399)	0.0129 (–0.00008)	0.6303 (0.3526)	0.6443 (0.3525)	0.0140 (–0.0001)	0.7405 (0.3570)	0.7482 (0.3570)	0.0077 (–0.00004)
Ghana	–0.3278	–0.1747 (0.3558)	–0.1550 (0.3553)	0.0197 (–0.0004)	–0.1407 (0.3642)	–0.1192 (0.3637)	0.0214 (–0.0005)	0.1138 (0.3568)	0.1256 (0.3566)	0.0118 (–0.0001)
Kenya	0.3421	0.7192 (0.3391)	0.7357 (0.3391)	0.0164 (–0.00001)	0.6200 (0.3467)	0.6379 (0.3467)	0.0178 (–0.00005)	0.7114 (0.3498)	0.7213 (0.3497)	0.0098 (–0.0001)
Madagascar	–0.2026	0.1932 (0.3402)	0.2085 (0.3401)	0.0152 (–0.00004)	0.1031 (0.3457)	0.1197 (0.3457)	0.0165 (0.00003)	0.2426 (0.3592)	0.2517 (0.3591)	0.0091 (–0.0001)
Malawi	0.5927	0.8230 (0.3513)	0.8303 (0.3515)	0.0072 (0.0002)	0.7242 (0.3564)	0.7320 (0.3566)	0.0078 (0.0002)	0.5100 (0.3597)	0.5144 (0.3599)	0.0043 (0.0002)
Mali	–0.0373	0.2486 (0.3381)	0.2615 (0.3383)	0.0129 (0.0001)	0.1460 (0.3487)	0.1600 (0.3487)	0.0140 (0.00006)	0.1680 (0.3528)	0.1757 (0.3528)	0.0077 (0.00005)

Mauritania	0.2896	0.5856	0.5985	0.0129	0.6351	0.6491	0.0140	0.6048	0.6126	0.0077
		(0.3411)	(0.3412)	(0.00002)	(0.3542)	(0.3542)	(–0.00005)	(0.3558)	(0.3559)	(0.0001)
Mauritius	0.4080	0.5776	0.5984	0.0207	0.5172	0.5398	0.0225	0.6561	0.6686	0.0124
		(0.3358)	(0.3353)	(–0.0004)	(0.3465)	(0.3460)	(–0.0005)	(0.3495)	(0.3490)	(–0.0005)
Niger	0.4449	0.2695	0.2825	0.0129	0.0921	0.1061	0.0140	0.4615	0.4693	0.0077
		(0.3470)	(0.3470)	(–0.00002)	(0.3517)	(0.3518)	(0.00006)	(0.3604)	(0.3604)	(–0.00006)
Nigeria	0.1170	–0.1091	–0.0823	0.0267	–0.2440	–0.2149	0.0290	0.1940	0.2101	0.0160
		(0.3563)	(0.3551)	(–0.0011)	(0.3681)	(0.3668)	(–0.0013)	(0.3493)	(0.3483)	(–0.0010)
Rwanda	0.4141	0.1590	0.1755	0.0164	0.0681	0.0860	0.0178	0.0223	0.0321	0.0098
		(0.3404)	(0.3404)	(0.00004)	(0.3534)	(0.3534)	(0.00003)	(0.3559)	(0.3559)	(–0.0001)
Senegal	0.0408	0.0983	0.1096	0.0113	0.1258	0.1381	0.0122	0.1546	0.1614	0.0068
		(0.3347)	(0.3347)	(0.00006)	(0.3441)	(0.3442)	(0.00007)	(0.3498)	(0.3499)	(0.0001)
Sierra Leone	0.4545	0.0947	0.1020	0.0073	0.4270	0.4350	0.0079	0.5964	0.6008	0.0043
		(0.3549)	(0.3550)	(0.0001)	(0.3505)	(0.3507)	(0.0002)	(0.3588)	(0.3590)	(0.0002)
Somalia	–0.3158	0.1451	0.1573	0.0122	0.2648	0.2781	0.0132	0.2932	0.3005	0.0073
		(0.3466)	(0.3466)	(0.00005)	(0.3524)	(0.3525)	(0.00005)	(0.3511)	(0.3512)	(0.0001)
South Africa	0.3931	0.4022	0.4151	0.0129	0.4877	0.5018	0.0140	0.3856	0.3934	0.0077
		(0.3382)	(0.3381)	(–0.0001)	(0.3589)	(0.3588)	(–0.0001)	(0.3555)	(0.3550)	(–0.0005)
Sudan	–0.1890	–0.419	–0.0221	0.0198	0.1926	0.2142	0.0215	0.1922	0.2041	0.0119
		(0.3417)	(0.3414)	(–0.0003)	(0.3504)	(0.3499)	(–0.0004)	(0.3484)	(0.3480)	(–0.0004)
Tanzania	0.6172	0.5078	0.5181	0.0103	0.5549	0.5661	0.0111	0.4754	0.4816	0.0061
		(0.3515)	(0.3517)	(0.0002)	(0.3594)	(0.3596)	(0.0002)	(0.3674)	(0.3676)	(0.0002)
Togo	0.2301	0.6298	0.6427	0.0129	0.6659	0.6799	0.0140	0.7188	0.7265	0.0077
		(0.3372)	(0.3374)	(0.0001)	(0.3462	(0.3463)	(0.00008)	(0.3508)	(0.3510)	(0.0002)
Uganda	0.1042	–0.3093	–0.3031	0.0061	–0.1412	–0.1345	0.0067	–0.2827	–0.2790	0.0037
		(0.3491)	(0.3493)	(0.0001)	(0.3568)	(0.3571)	(0.0003)	(0.3678)	(0.3681)	(0.0003)
Zaire	–0.3659	0.0333	0.0406	0.0073	0.1260	0.1339	0.0079	0.3452	0.3496	0.0043
		(0.3487)	(0.3489)	(0.0001)	(0.3599)	(0.3602)	(0.0003)	(0.3565)	(0.3568)	(0.0003)
Zambia	–0.1472	0.2496	0.2605	0.0109	0.6691	0.6810	0.0119	0.6025	0.6091	0.0065
		(0.3575)	(0.3576)	(0.00003)	(0.3588)	(0.3589)	(0.0001)	(0.3578)	(0.3579)	(0.0001)
Zimbabwe	0.5738	0.8728	0.8865	0.0137	0.6666	0.6815	0.0149	0.6326	0.6408	0.0082
		(0.3374)	(0.3374)	(0.00005)	0.3462)	(0.3462)	(0.00007)	(0.3465)	(0.3467)	(0.0002)

Source: Authors' regressions.

a. The mean and variance of the growth rate for each country between 1960 and 1985 are compared with estimates of those that would have occurred with a 10 percent reduction in tariffs. Posterior means and variances of model parameters β, γ, and δ are taken from historical data. The posterior mean and variance of g_i are then computed with and without a 10 percent reduction in the tariff variable, keeping all other regressor values constant. Errors in the growth process are assumed to be independent of the regressors. Theory inclusion probability q is 0.50 in all regressions; heterogeneity is not considered. Standard deviations are shown in parentheses.

that matter on the other regressors). 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 that gave the largest and the smallest tariff coefficients. For the model averaging exercises, we employ a theory inclusion probability $q = 0.5$, which reflects our judgment that the theories we have allowed for are quite plausible ex ante, that is, 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 probability is 0.5 or 0.75. In addition to the model averaging exercises, table 4 reports results for the models with the largest and the smallest tariff parameters.

What sorts of conclusions might one draw from the information in table 4? One finding of importance is the heterogeneity in expected growth levels across countries. To focus on the estimates under model averaging, Botswana, for example, is associated with expected growth over this period of over 100 percent (under the historic level of the tariff variable), whereas Burundi's expected growth was –9 percent. The differences in the standard deviations are much smaller, because 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 equation 45 is not general. It is easy to construct loss functions that would lead one to 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 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 table 4.

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 by 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.[94] 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 that applies to the rest of the world does not apply to that region.

Conclusions and Suggestions for Future Research

In this paper we have attempted to exposit a perspective on policy evaluation that explicitly places such evaluation in a decision-theoretic context and that 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 offered some suggestions about how to implement this approach empirically. An example of empirical implementation to growth

94. Brainard (1967).

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 obtains 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 of models that a policymaker perceives as possible descriptions of the economy evolves over time; as we 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 suggests that the macroeconomy exhibits shifts across regimes;[95] 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 to illustrate how to calculate policy effects while 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-

95. Hamilton (1989).

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.

APPENDIX A

Methodology

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

(A1) $$g_i = S_i\zeta + \varepsilon_i \quad i = 1 \ldots I.$$

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

(A2) $$\mu(\zeta) \propto c.$$

Second, we assume that the errors are i.i.d. (identically and independently distributed) normal with a known variance. Under this assumption, one can show that the posterior density of the regression coefficients is[96]

(A3) $$\mu\left(\zeta \middle| d, m\right) = N\left[\hat{\zeta}, (S'S)^{-1}\sigma_\varepsilon^2\right],$$

where $\hat{\zeta}$ is the OLS estimate of the coefficient parameters in equation A1. 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

96. Box and Tiao (1973, p. 115).

interpretations. The assumption that the error variance is known is not a serious problem when the number of observations is large relative to the number of regressors.

A considerable discussion in the literature concerns the appropriate choice of priors even for this model. Fernández, Ley, and Steel consider a range of alternative priors and argue in favor of a different set of priors than those we employ.[97] 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

Monetary Policy

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

Growth

All Bayesian model averaging exercises in the growth section of the paper were calculated using the SPLUS statistical package. 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 at the University of Washington and available on the Internet at lib.stat.cmu.edu/S/bicreg. This procedure uses an "Occam's window" procedure due to David Madigan and Raftery.[98] In adapting the code for our exercise, a few adjustments were necessary and are available from the authors.

Prior probabilities were set as follows. For a given growth specification, one first specifies the probability that a given theory is included. Table 2 allows these probabilities to be 0.25, 0.5, and 0.75. For a given theory, with r empirical proxies, there are $2^r - 1$ different ways to include

97. Fernández, Ley, and Steel (2001a).
98. Madigan and Raftery (1994).

these proxies. All specifications are 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. To handle these calculations, we follow an approximation suggested by Raftery,[99] which exploits the fact that 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 Bayesian information criterion-adjusted likelihoods. This greatly simplifies the calculation of posterior model probabilities.[100] Of course, the approximation becomes more accurate the larger the data set. The program for this approximation is taken from *bicreg* described above.

APPENDIX B

Data Definitions and Sources

Monetary Policy

All data in the monetary policy section 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 GDP price index. Potential GDP is the Congressional Budget Office measure. The quarterly average federal funds rate was computed by averaging monthly average figures.

99. Raftery (1995).

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

Growth

Table B1 lists the various growth variables, their definitions, and their sources.

Table B1. Variables Used in the Growth Regressions

Variable	*Definition*	*Source*
Solow variables[a]		
MNGD	Log($n + g + d$), where n is the rate of population growth, g the exogenous rate of technical change, and d the rate of depreciation; $g + d$ is assumed to equal 0.05 for all countries.	Mankiw, Romer, and Weil (1992)
MINV	Logarithm of the investment rate	Mankiw, Romer, and Weil (1992)
MSCH	Logarithm of the fraction of the population aged 12–17 enrolled in school multiplied by the fraction of the working-age population aged 15–19	Mankiw, Romer, and Weil (1992)
MGDP60	Logarithm of income per capita in 1960	Mankiw, Romer, and Weil (1992)
Policy variables		
Tariffs		
OWTI	Country's own import-weighted tariff rate on intermediate and capital goods	Barro and Lee (1994)
Exchange rates		
BMPL6089	Logarithm of the average black-market premium on the domestic currency, 1960–89, log(1 + BMP)	Barro and Lee (1994)
RERD	Measure of real exchange rate distortion	Dollar (1992)
Inflation		
PI6089	Average annual inflation rate for 1960–89	Sala-i-Martin, citing data in Levine and Renelt (1992)
PIHYP6089	Dummy variable equal to 1 when PI6089 > 15 percent	Calculated by the authors from data for PI6089
Government spending		
GGCFD	Ratio of real public domestic investment to real GDP	Barro and Lee (1994)
GVXDXE5	Ratio of real government "consumption" spending (net of spending on defense and education) to real GDP	Barro and Lee (1994)

(*continued*)

Table B1. Variables Used in the Growth Regressions (*continued*)

Variable	*Definition*	*Source*
Structural variables		
Economic structure		
EcOrg	Index of the degree of capitalism prevailing in the country as measured by Freedom House	Sala-i-Martin (1997)
RULELAW	Index of the rule of law prevailing in the country	Sala-i-Martin (1997)
Financial structure		
DCPY	Ratio of gross claims on the nonfinancial private sector by the central bank and deposit banks to GDP	King and Levine (1993)
LLY	Ratio of liquid liabilities of the financial system to GDP	King and Levine (1993)
Political structure		
CIVILLY	Index of civil liberties	Knack and Keefer (1995)
DMCYBL	Index of democracy, ranging from 0 to 1, where 1 = most democratic	Barro and Lee (1994)

a. A constant term is included as a Solow regressor in all specifications.

Comments and Discussion

Eric M. Leeper: This is an unusual paper for the Brookings Panel because it is heavily methodological. I mention this at the outset, not to criticize the paper or Brookings, but to urge people who are serious about giving policy advice to read and study this paper, despite its methodological bent. This is a provocative paper: it provoked me to rethink how I would formulate and present policy advice. Any policy adviser who digests the paper's central messages will give better and more pertinent advice. In this comment I adopt the perspective of a policy adviser—a perspective from which it is apparent where the paper's marginal product lies.

This comment consists of two parts. The first part reviews the substantial progress that William Brock, Steven Durlauf, and Kenneth West have made toward embedding model uncertainty and policy evaluation in a decision-theoretic framework. I highlight those aspects of their work that develop methods of reporting results that differ from the norm in ways that can be helpful to policymakers. The second part points out the ways in which the authors' current development of their methods falls short of offering a framework for policy evaluation that can be taken immediately into the briefing room. Before their methods can speak to actual policy questions, they need to be extended to confront problems of identification and to incorporate dynamics along with the kinds of strategic interactions that arise in rational expectations equilibria. I conclude with a discussion of some practical problems associated with communicating policy analysis that accounts for model uncertainty.

The approach that Brock, Durlauf, and West develop is designed to answer exactly the right question: how can we design good policies in the face of extreme uncertainty about how the economy works? Model uncertainty, almost everyone agrees, is important, yet it is grossly understudied. And accounting for this kind of fundamental uncertainty is crucial for policy analysis when the social costs of basing decisions on a bad model can be very high.

One appealing aspect of this approach, which the authors do not emphasize, is that it formalizes and systematizes an informal and unsystematic process that already takes place in policy institutions. Current practice in policy analysis brings many disparate models to bear on the questions at hand. At Federal Open Market Committee (FOMC) meetings, for example, there are probably at least as many models and priors over models as there are participants. Models are combined with data in both coherent and incoherent ways. Acknowledgment of model uncertainty and a type of model averaging both take place during policy debates, as arguments frequently draw on implications from different models. Because all of this occurs informally, there is little basis for comparing competing models. And although the models are presented as competing, the rules of the competition are only vaguely spelled out. At the end of the day, the "winner" is typically the model on which the policymakers base their decisions, rather than the model that most accurately represents the economy. Unfortunately, there is no assurance that models demonstrated to be at odds with the data will be discredited and eventually disappear.[1] The authors offer a cure for this common practice by creating a framework for rigorous discussion of alternative models.

It may be especially difficult to apply the authors' methodology to U.S. monetary policy. It is unlikely that the governors of the Federal Reserve System and the presidents of the regional Federal Reserve banks could easily agree on a single loss function for the Fed.[2] But monetary policy analysis in countries that announce an explicit inflation target may readily lend itself to the authors' approach. Indeed, by adopting their approach and explaining it in their periodic inflation reports, inflation-targeting central banks could derive the benefits that a more formal and systematic approach to policymaking offers. That approach may also carry with it

1. For example, the notion that further reductions in tax rates will generate a burst of supply-side activity that raises sufficient revenue to pay for the tax reduction is just as alive in some circles today as it was over twenty years ago, when the federal budget deficit skyrocketed.

2. Of course, an individual regional Federal Reserve bank with a single decisionmaker could implement the approach.

some positive externalities, in the form of greater accountability and transparency.

Reflecting on how actual policymaking is practiced, with its explicit, although typically informal, recognition of model uncertainty, one is struck by how wide the gulf is between this practice and research on policy evaluation. It is fashionable to pose optimal monetary and fiscal policy questions as Ramsey problems, which solve for the policies that select the best competitive equilibrium. Although it is understood that the answer to the question of which policy is optimal is strongly model dependent, varying both with the frictions present in the models and with the policy instruments assumed to be available,[3] no analysis of optimal policy proceeds by first averaging across models according to their fit to data. The present paper has the potential to bridge that gulf and bring research and practice closer together.

The paper considers a potentially very large and disparate range of models and applies formal statistical evaluation to them. Models unsupported either by the policy advisers on a priori grounds or by the data are given little weight in the posterior density. The authors embed this statistical analysis in a decision-theoretic framework for policy evaluation as a means of arriving at optimal policy rules in uncertain economic environments.

The paper presents two extended empirical examples. The first is the "backward-looking," reduced-form monetary model of Glenn Rudebusch and Lars Svensson,[4] for which the authors compute the mapping from weights in the policy loss function to parameters of the optimal Taylor rule, after averaging across 25,600 variants of the model. Many of their findings are close to Rudebusch and Svensson's. Assuming that Rudebusch and Svensson pretested their specification, the similarity of results is not too surprising given that the present authors' procedure assigns small weights to ill-fitting models.

This example does not do justice to the richness of the authors' approach, however. The variants of the models in the example are really quite close to each other and hard to distinguish empirically. Moreover, whether three or four lags of output enter the IS curve is not the kind of uncertainty that makes policy discussions heated. Nor is it the type of

3. See, for example, Schmitt-Grohé and Uribe (2000, 2002, forthcoming).
4. Rudebusch and Svensson (1999).

uncertainty that is likely to lead to very bad policy choices based on the wrong model.

Instead, the uncertainty that matters arises when, for example, one adviser points to low inflation figures and a federal funds rate of 1 percent to underscore worries about deflation, while another cites four consecutive quarters of rapid M2 growth to argue that deflation is not even a remote concern. The models behind each piece of advice differ dramatically—probably by more than any two of the 25,600 models the authors consider. But their methodology can in principle be applied to models that differ greatly.

The authors' second example uses cross-country growth regressions to address the question of how tariffs affect growth. The authors consider three levels of uncertainty—different growth theories, different empirical proxies for the key variables in the theories, and different assumptions about the heterogeneity of countries' growth processes—in a total of 8,192 models. Even someone who regards growth regressions as reduced-form specifications that cannot offer clear policy advice would find this analysis fascinating. The authors show how policymakers with various kinds of preferences would interpret and act on very different statistics. An adviser must be sufficiently attuned to the policymaker's preferences to present evidence that speaks to the policymaker's concerns. The example also illustrates how an advocate of a particular policy choice can mine the data to find evidence to persuade policymakers of that choice.

These are important insights, and they are applicable to the current monetary policy environment. To explain its May 2003 decision to leave the federal funds rate unchanged, the FOMC said in its public statement that ". . . the probability of an unwelcome substantial fall in inflation, though minor, exceeds that of a pickup in inflation from its already low level."[5] Even though FOMC members claim that deflation is extremely unlikely, its ill effects are deemed sufficiently great that policymakers adopted an asymmetric policy directive. This is a case where merely reporting central tendencies, as policy advisers are wont to do, simply would not address the policymakers' concerns.

Let me now turn to some possible extensions of the authors' approach. What this paper offers is a first step toward a practical framework for

5. Federal Open Market Committee press release, May 6, 2003 (federalreserve.gov/boarddocs/press/monetary/2003/20030506/default.htm).

policy analysis. Before I would know how to take the framework into a policy briefing, some further problems need to be worked out.

The first necessary extension involves identification. Both the examples that the authors present are reduced-form setups in which it is unclear how to interpret the model-averaged results in terms of economic behavior. It would be instructive to work out an example in which uncertainty is concentrated in a set of "deep" parameters, π, describing preferences or technologies. The prior distribution, $p(\pi)$, over those parameters would also represent the prior distribution over economic models. Reduced forms would be indexed by π. One could then proceed with model averaging and estimation to obtain the posterior density function. Model-averaged results would then be clearly interpretable, because the posterior distribution for π would be connected to well-defined economic behavior.

In this more detailed description of private behavior, one might want to take a more symmetric position on the treatment of uncertainty. In the present paper, policymakers are ignorant of the "true" model, whereas private agents happily inhabit the truth. In a more symmetric treatment, private agents might understand their local environment but be uncertain about the aggregate laws of motion. At the same time, the policy authorities would entertain a wide set of possible models, just as the authors imagine.

A second important extension involves dynamics. Extending the method to incorporate dynamics would serve several purposes. First, it would allow policy evaluation to confront the Lucas critique head on. This is generally important, but it seems particularly so for the kinds of once-for-all policy choices discussed in this paper.

Second, dynamics allows the modeling of learning, by both the private sector and the policy authority. Here two possibilities offer themselves. On the one hand, it would seem that uncertainty might become less diffuse over time, as additional data alter the posterior probabilities and, therefore, policy rules and private decision rules. On the other hand, there does not appear to be much evidence that this kind of convergence on models actually takes place. The introduction of model innovation would create a dynamic that prevents convergence and may generate interesting dynamics in policy choices.

Third, dynamics might lead to a description of the authors' approach for ongoing policy analysis. Over time, as new policy problems arise and the economy changes, the set of relevant models will also change. How can their approach evolve as, over time, policymakers apply it and private agents react to its application?

Because they have focused on developing their methods, the authors naturally did not confront some practical issues surrounding how to get these methods into policy meetings. Several steps are involved in actually using their methods. First, policymakers must buy into the notion of formalizing the analysis of model uncertainty. Although policy institutions readily admit there is no single, universally accepted model of the economy, there is little evidence that those institutions embrace the idea of juggling many, possibly quite different, representations of reality. The closest central banks come to this is the presentation of alternative scenarios in briefing materials. But these really represent alternative realizations of exogenous shocks, or alternative paths of policy instruments, rather than predictions from alternative economic structures. These alternative scenarios capture one or two degrees of uncertainty, but not the fundamental uncertainty that concerns the authors.

Second, a policy adviser must get inside the policymaker's head, to try to discern the relevant loss function. This is a difficult task, because many policymakers are reluctant to reveal their preferences or are simply unable to articulate them. But without a clear understanding of the loss function, the adviser cannot effectively address pertinent issues and present useful analysis.

Third, there is the tricky question of how to present model-averaged results. Story telling is a key aspect of policy advice. Compelling stories get retold by policymakers when they argue their viewpoint before other policymakers and the public. Model uncertainty muddies the waters and can make the story underlying a policy recommendation murky and less compelling. At present, I fear, the authors' approach is too much of a "black box" for policymakers to find it palatable.

Although the authors have impressively shown that their approach can handle a huge number of alternative models, for practical policy analysis it is not obvious that this is the most productive way to proceed. Differences of opinion about the appropriate model usually concentrate on a small handful of alternative structures. By narrowing the class of structures, the adviser can focus discussion on the fit and the implications of each viewpoint, in the hope of narrowing the differences still more.

Thomas J. Sargent: Appealing to statistical decision theories, William Brock, Steven Durlauf, and Kenneth West recommend that econometric analysis be thoroughly integrated with the policy evaluation process in new ways. This is a two-way street. First, if it is to be used to evaluate policy, econometric work should not be directed toward model selection tests, because the small samples that economists typically must work with do not allow one to choose among competing specifications with much confidence. Instead, econometric work should strive accurately and forcefully to present measures of doubt that properly account for the components of ignorance due to ambivalence about alternative plausible specifications. Second, quantitative policy evaluation exercises should take account not only of parameter uncertainty but also of uncertainty about model specifications. Thus a set of alternative specifications should be on the table when policy proposals are analyzed. In what is a popular mainstream macroeconomic tradition nowadays, researchers first "calibrate," then solve for the optimal policy for a given set of parameters. In such exercises the formal procedures that the authors advocate are not being used, although limited forms of robustness analysis that check the sensitivity of results to parameter values are often performed.

What, precisely, does one mean by model uncertainty? The authors do a good job of formalizing two distinct possible meanings. The first is methodologically the more conservative, because it is amenable to business-as-usual Bayesian modeling. According to this definition, the decisionmaker has in mind a sufficiently small set of alternative specifications to assign meaningful prior probabilities to them. After that prior is taken into account, there really only remains a single model on the table (remember that a model is a probability distribution over sequences of outcomes, that is, a likelihood function). As Ramon Marimon puts it,[1] a Bayesian decisionmaker acts as if he or she knows the truth (that is, a unique model) from day one.

But how should the decisionmaker proceed if he or she cannot put a prior over a given set of models or, if forced to do so, can only name a set or range of priors? And what if the decisionmaker cannot articulate a set of models but instead can only describe one (or at most a small number), but nevertheless views that model as an approximation to some unknown data-generating mechanism that the decisionmaker cannot specify? These

1. Marimon (1997).

types of uncertainty about models do not fit neatly into the Bayesian single-model framework. They present the decisionmaker with the need somehow to act while acknowledging multiple possible models. Taking seriously the notion that a model is an approximation (to what?) thus seems to require leaving the Bayesian setting and embarking on a quest for decision rules that are robust across a range of model specifications. "Robust" rules here mean rules that "work well enough," that is, that assure some minimal acceptable level of performance regardless of the data-generating mechanism.

To investigate the performance of a decision across a set of models, we are impelled to compute *bounds.* A good tool for doing that is, for a given decision rule, to study the consequences of *minimizing* the decisionmaker's objective function by pretending that one can choose the data-generating mechanism from among the set of models under consideration. It happens that robust decision rules can be attained using a minimax rule. Confronted with model uncertainty, the decisionmaker invents a hypothetical malevolent agent to guide the analysis of the fragility of a proposed decision rule to model misspecification.

The authors give some simple examples that illustrate the types of analysis they recommend. Partly reflecting their roots in classical statistical decision theory, their examples share two features: they refer to a single decisionmaker facing an environment that is not explicitly populated by other decisionmakers, and the decision problems and the competing models are static. But to appreciate the rich substantial implications of the methodological reforms that the authors advocate for existing macroeconomic theory and practice, one should look at multiple-agent dynamic models.

When one begins to think about misspecification in the context of dynamic macroeconomic models with multiple agents, one is immediately forced to reconsider rational expectations. For thirty years rational expectations has been used to economize ruthlessly, like a communist, on the number of models that in a multiple-agent setting must necessarily reside within the model itself (because the decisionmakers themselves have beliefs, that is, models). I say "like a communist" because, in a rational expectations model, the diverse agents being modeled, the theorist, and the econometrician are all forced by the theorist to share the same model, that is, the same probability distribution over the sequence of outcomes. They may have different information, but they agree about the model. Within a rational expectations model, decisionmakers neither acknowledge nor fear model misspecification, nor should they. The uniqueness of the model eliminates decisionmakers' beliefs as free parameters (instead they become outcomes of the analysis) and is the source of the powerful cross-equation restrictions exploited by rational expectations econometrics.

That uniqueness of models vanishes in the presence of multiple decisionmakers who acknowledge model uncertainty. The challenge will be to preserve some of the empirical power of rational expectations econometrics in this situation. For example, in a macroeconomic model designed to analyze fiscal policy, to which decisionmakers do we choose to attribute model uncertainty? The government, understood as a Ramsey planner? some of the consumers and firms within the model? everybody? And if we choose to characterize model uncertainty by attributing *sets* of models to different decisionmakers, how should we characterize each decisionmaker's knowledge about the sets of models that concern other decisionmakers? Researchers have only begun to think about these challenging questions about modeling strategy. One approach has been to attribute concerns about model misspecification only to the government decisionmaker, and to regard the private agents as knowing a model.[2]

In an effort to retain as many of the advantages of rational expectations as we can, Lars Hansen and I have taken another approach that we think appropriately generalizes rational expectations to acknowledge widespread model uncertainty.[3] We impose on all agents a common approximating model, but we allow each decisionmaker to have his or her own set of alternative models around that model.[4] The common approximating model has to incorporate the robust decision rules of the various decisionmakers who compose the economy.[5] Although they share a common

2. For example, see Giannoni (2002). In this regard, Giannoni's analysis shares some features with Otrok's (2001) Bayesian econometric analysis of a real business cycle model. In Otrok's work the econometrician explores the likelihood function for different vectors of parameter values, while assuming that the agents inside the model *know* each such candidate vector of parameter values.

3. Hansen and Sargent (2002c).

4. One can reinterpret the work of Anderson (2003) on multiple-agent, risk-sensitive economies as embracing a common approximating model.

5. This is a natural setting for thinking about some comments about model uncertainty that Federal Reserve Board Governor Ben Bernanke gave at a conference on learning at the

approximating model, the fact that decisionmakers have different preferences means that their individual minimax calculations yield *different* context-specific worst-case models. These differences partly reflect the fact that, because their preferences differ, the various decisionmakers' rules are fragile to different departures from the common approximating model.

Even for a single decisionmaker, dynamic settings introduce significant complications into the task of modeling uncertainty about model misspecification in terms of sets of models. Some of the complications pertain to technical issues about how the minimax decisionmaking of the static model transfers to a dynamic setting, and in particular to whether the minimization and maximization can each be decentralized over time so that the decision problem can be rendered recursive. Here we are in luck, because the dynamic decision problem is a zero-sum, two-player dynamic game with substantial structure that, under a technical Bellman-Isaacs condition, renders the minimax problem recursive.

In dynamic applications it can be especially interesting to inspect the worst-case model that accompanies a robust decision analysis. For a permanent income model, Hansen, Thomas Tallarini, and I show that the consequence of a consumer's concern about possible misspecification of the stochastic process governing the components of his or her endowment process is to generate a form of precautionary saving, because the most troublesome and hard-to-self-insure-against misspecifications are at the low-frequency components.[6] The minimizing player cannot harm the consumer by tampering with high-frequency components, because the permanent income model does such a good job of smoothing those components. Therefore the minimizing agent puts the worst-case misspecification at low frequencies, and the consumer recognizes and responds to this strategy by saving more.

The unique model presumed by the Bayesian paradigm gives rise to a complete and convenient theory of learning based on Bayes' Law. Models of model uncertainty that have the decisionmakers entertain multiple models require modifying the Bayesian theory of learning. An active

---

Federal Reserve Bank of Atlanta on March 22, 2003 (remarks not published). Bernanke spoke about downstream effects of risk and model uncertainty. He pointed out that even decisionmakers who are confident of their model specification are wise to respond to the price and quantity signals that reflect the model uncertainties of *other* decisionmakers.

6. Hansen, Sargent, and Tallarini (1999).

area of research investigates how to model learning in the presence of multiple models, an issue that interacts in interesting ways with the time-consistency issues just mentioned.[7]

Before concluding, I offer a digression about another kind of misspecification, one for which we need better theorists, not better statistical decision theories. The theory of self-confirming equilibria directs our attention to a type of model misspecification that even unlimited amounts of historical statistical evidence are powerless to correct. The problem is that there are multiple observationally equivalent models, each of which can have different policy implications. Essentially, these models have identical likelihood functions, not over all conceivable events, but conditional on events that occur "infinitely often" in equilibrium.[8] Many important disputes in macroeconomics have been about which of such observationally equivalent models is better. One example is the dispute over various expectational and nonexpectational Phillips curves at the dawn of the process of bringing rational expectations into macroeconomics. A second involves the choices about how to form vectors of interpretable shocks from the correlated innovations recovered by a vector autoregression (this is the process of creating a just identified vector autoregression).

To conclude, the authors are tackling a difficult set of problems, which over the years have challenged and defeated some brilliant minds. Their quotation from Keynes is apt. It alludes to what is either a deep insight or a theoretical pipe dream and sideshow, depending on whom you ask. Keynes's *Treatise on Probability* and Frank Knight's *Risk, Uncertainty, and Profit* told us that decisionmakers often must act in the face of a kind of ignorance that cannot be described by the usual laws of probability; Keynes and some other smart people told us that this is a big deal for macroeconomics.[9] But taking cues from Frank Ramsey's paper "Truth and Probability," Leonard Savage elegantly proved Keynes wrong by completing a line of work that culminated in Savage's model of a rational decisionmaker as a Bayesian.[10] There is no sense in which Savage's

7. Important and interesting work is being done on this question by Epstein and Schneider (2002) and Knox (2002), among others.

8. Here I interpret "in equilibrium" as "historically." See Sargent (1999) for a discussion of self-confirming equilibria in macroeconomics.

9. Keynes (1921); Knight (1921).

10. Savage (1954).

decisionmaker fears model misspecification. But, as the authors emphasize, it may be that Daniel Ellsberg, and Itzhak Gilboa and David Schmeidler,[11] have taken the high ground from the Bayesians and put fear of model misspecification at the center of statistical decisionmaking.

For macroeconomists of the last thirty years, these are unfamiliar and challenging issues to think about. In our econometric practice and model building, most of us have followed John Muth in equating objective and subjective distributions, thereby conveniently eradicating any grounds for controversy between "frequentist" and "subjectivist" interpretations of probability, to say nothing of the more extensive distinctions that the authors have asked us to consider.

General discussion: Christopher Sims took issue with Thomas Sargent's suggestion that the Bayesian approach had lost the high ground. Sargent had mentioned two attacks on Savage's approach to decision theory: the Ellsberg paradox and the axiomatic system of Gilboa and Schmeidler. Sims viewed the Ellsberg paradox as a failure of perfect computing by people, not of the Bayesian description of optimal behavior. Following the Bayesian program for decisionmaking is hard, and frequently it cannot be followed exactly; people may react differently from what Bayesian theory prescribes. The descriptive implication is that economists should not assume that agents act like Bayesians with infinite computing capacity. But with respect to normative behavior, Sims argued that a decisionmaker should strive to achieve the Bayesian goal, even if it is too difficult to execute all Bayesian computations in real time.

Sims noted that if the explanation of the Ellsberg paradox is limited computing power, it does not suggest fundamental differences between Savage's approach and Knightian uncertainty. As a practical matter, it is sometimes too hard to create a formal model that incorporates all sources of uncertainty. However, this does not mean that there are two kinds of uncertainty in the world, nor does it mean that the Bayesian approach cannot in principle encompass Knightian uncertainty. Even where formulating a complete prior is impossible, it may be easy to see that a "robust" analysis produces a result that can be justified by no prior, or only by priors that are plainly unreasonable. In such a case, we should all agree that the robust algorithm used to derive the result provides no basis for

11. Gilboa and Schmeidler (1989).

clinging to it. Sims believed that the justification for robust analyses, minimax approaches, and non-Bayesian approaches is limited computing power, and not the existence of some—in his view not well defined—concept of Knightian uncertainty that the Bayesian approach will never be able to analyze.

Thomas Sargent strongly disagreed with Sims on the interpretation of the Ellsberg paradox. In Ellsberg's experiment, under all priors but one, a participant should have preferred the urn with the uncertain distribution, and under the remaining prior the participant should have been indifferent between the two urns. This choice problem should have been easy to solve under Savage's assumptions. Savage himself participated in the experiment and did not behave as a Bayesian. Sargent also disagreed with Sims' view that Knightian uncertainty is not well defined. He noted that Gilboa and Schmeidler provide a rigorous axiomatic description basis for Knightian uncertainty. Sargent guessed that, without knowing the implications, Sims would consider the axioms put forward by Gilboa and Schmeidler as just as satisfactory a description of a rational person's behavior as Savage's. Although Gilboa and Schmeidler's axioms are not any more artificial than Savage's, they are able to generate the behavior seen in the Ellsberg experiments and justify a decisionmaker behaving as a minimax optimizer.

Sims responded that he was familiar with the Gilboa-Schmeidler axioms and did not regard them as a basis for Knightian uncertainty. They lead to decisions that can always be justified as based on some probabilistic beliefs. They are therefore not completely unreasonable, yet they are always subject to critique if they lead to decisions that can be justified only by evidently unreasonable probabilistic beliefs.

Sims stressed that, under the assumption of a finite model space, it is crucial whether the true model is included in this space or not. The authors argue that this is not important, since the analysis will eventually put a probability of 1 on the best-fitting model. But the model in a collection of false models that fits best by a likelihood criterion need not lead to the best decisions or to the smallest losses available from the collection of models. These points are made in detail in work by Frank Schorfheide of the University of Pennsylvania.

Sims also criticized labeling the pseudo-Bayesian approach as "frequentist." A frequentist insists that only the data be treated as random and given probabilities. The authors label their procedure frequentist even though it also puts probabilities on models, given the data. Sims suggested speaking of a "likelihood" perspective rather than a "frequentist" perspective.

Sims elaborated on the selection of priors for models. The paper does not consider countable infinite sequences of models, which actually underlie a lot of practical procedures, for example selecting the number of lags to be used in dynamic models. If the data gave enough evidence in favor of very long lags, models with long lags would be used, but generally models with fewer lags are used unless the data force one to do otherwise. This procedure can be rationalized by thinking of it as working with countable collections of models with different numbers of lags, and with a prior probability distribution over the collection of models. The Bayesian approach then rationalizes the procedure of working with smaller models unless the data strongly favor larger ones.

Sims questioned the proposed principle that a priori, for scientific reporting, one should always assign high probability to regions where the likelihood function is large. He illustrated this point by citing the examples of seemingly unrelated regressions and simultaneous equations models; in such models the likelihood function can be unbounded from above. When the number of equations is large, even when the usual measure of "degrees of freedom" is substantial, there may be an unbounded peak in the likelihood associated with a reduced-form residual covariance matrix that is singular. In this situation the likelihood will be well behaved around the finite peak near the ordinary least squares estimates of the reduced form, and the infinite peak is not interesting. In these situations a prior that assigns zero weight to the subset of the parameter space defined by the singularities is usually a good idea, even though this rules out a region with high likelihood.

Sims found the other two rules for selecting priors cited by the authors to be very reasonable for the scientific reporting problem. One should keep in mind that the econometrician is usually not making decisions directly, but rather reporting results of data analysis to an audience of interested people with differing views. The role of the prior is then only to assist in reporting the likelihood shape. Different readers of the report will apply their own priors in reaching decisions. However, if an econometrician reports directly to a decisionmaker who has to make a decision immediately, the principles associated with picking a prior should change. In this case the priors should reflect beliefs from all available sources of information, including those outside the observed data.

William Brainard applauded the authors' adoption of a decision-theoretic approach that recognizes the wide variety of candidate models and emphasizes the need to focus on the probability distribution of outcomes conditional on policy rather than only on statistical testing per se. But he felt that the usefulness of model averaging depended on the context. A policy recommendation based on performance averaged over a large number of models is not likely to be persuasive to a policymaker, nor are summary statistics about the performance of different policies across models sufficient. Policymakers have their own priors about the plausibility of models. In such situations it is important that the policymaker have some understanding of the alternative ways of looking at a problem and their differing implications for policy, as well as some sense of what the data suggest as their relative likelihoods. One needs to present a limited hierarchy of models capturing the most important alternative views of the world rather than a single policy recommendation based on performance over a large number of models.

Michael Woodford also congratulated the authors on addressing an important problem and making considerable progress toward solving it. However, he agreed with Eric Leeper on the need to move beyond the static concept of policy choice. In the monetary policy example, the role of the policymaker is to choose the coefficients for the Taylor rule once and for all. Hence the analysis relies on asymptotic criteria. Short-term objectives are given relatively little consideration; rather the goal is to minimize the variability of output, inflation, and interest rates on the assumption that the chosen system is run forever. However, policymakers, appropriately, are not worried that a policy might produce instability in the long run, because adjustments are always possible if they become necessary. It is nevertheless not trivial to distinguish between risks that can be easily addressed in the long run, and policies that can be corrected only with considerable costs. Woodford suggested that the authors investigate this central issue further.

References

Anderson, Evan W. 2003. "The Dynamics of Risk Sensitive Allocations." University of North Carolina.

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

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

Barro, Robert J., and Jong-Wha Lee. 1994. "Data Set for a Panel of 138 Countries." Cambridge, Massachusetts: National Bureau of Economic Research. Available at www.nber.org/pub/barro.lee.

Barro, Robert J., and Xavier Sala-i-Martin. 1995. *Economic Growth*. New York: McGraw-Hill.

Becker, Selwyn W., and Fred O. Brownson. 1964. "What Price Ambiguity? or the Role of Ambiguity in Decision-Making." *Journal of Political Economy* 72(1): 62–73.

Berger, James O. 1985. *Statistical Decision Theory and Bayesian Analysis*. New York: Springer-Verlag.

Bernard, Andrew B., and Steven N. Durlauf. 1996. "Interpreting Tests of the Convergence Hypothesis." *Journal of Econometrics* 71(1–2): 161–73.

Bernardo, José M., and Adrian F. M. Smith. 1994. *Bayesian Theory*. New York: John Wiley and Sons.

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

Blinder, Alan S. 1997. "Distinguished Lecture on Economics in Government: What Central Bankers Could Learn From Academics—and Vice Versa." *Journal of Economic Perspectives* 11(2): 3–19.

Box, George E. P., and George C. Tiao. 1973. *Bayesian Inference in Statistical Analysis*. Reading, Mass.: Addison-Wesley.

Brainard, William C. 1967. "Uncertainty and the Effectiveness of Policy." *American Economic Review* 57(2): 411–25.

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

———. Forthcoming. "Design Limits Analysis and Local Robustification Analysis: Theory and Applications to Economics." University of Wisconsin.

Brown, Philip J., Marina Vannucci, and Tom Fearn. 1998. "Multivariate Bayesian Variable Selection and Prediction." *Journal of the Royal Statistical Society,* series B 60(3): 627–41.

———. 2002. "Bayes Model Averaging with Selection of Regressors." *Journal of the Royal Statistical Society,* series B 64(3): 519–36.

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

Canova, Fabio. 1999. "Testing for Convergence Clubs in Cross-Country Growth Data: A Predictive Density Approach." CEPR Discussion Papers 2201. London: Centre for Economic Policy Research.

Chamberlain, Gary. 2001. "Econometrics and Decision Theory." *Journal of Econometrics* 95(2): 255–83.

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

Cox, David R., and David V. Hinkley. 1974. *Theoretical Statistics*. London: Chapman and Hall.

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

Dehejia, Rajeev H. Forthcoming. "Program Evaluation as a Decision Problem." *Journal of Econometrics.*

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

Dollar, David. 1992. "Outward-Oriented Developing Economies Really Do Grow More Rapidly: Evidence from 95 LDC's, 1976–1985." *Economic Development and Cultural Change* 40(3): 523–44.

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

Draper, David. 1995. "Assessment and Propagation of Model Uncertainty." *Journal of the Royal Statistical Society,* series B 57(1): 45–70.

Durlauf, Steven N. 2000. "Econometric Analysis and the Study of Economic Growth: A Skeptical Perspective." In *Macroeconomics and the Real World,* vol. 1*: Econometric Techniques and Macroeconomics*, edited by Roger E. Backhouse and Andrea Salanti. Oxford University Press.

Durlauf, Steven N., and Paul A. Johnson. 1995. "Multiple Regimes and Cross-Country Growth Behaviour." *Journal of Applied Econometrics* 10(4): 365–84.

Durlauf, Steven N., Andros Kourtellos, and Artur Minkin. 2001. "The Local Solow Growth Model." *European Economic Review* 45(4–6): 928–40.

Durlauf, Steven N., and Danny T. Quah. 1999. "The New Empirics of Economic Growth." In *Handbook of Macroeconomics*, vol. 1a, edited by John Taylor and Michael Woodford. Amsterdam: North Holland.

Easterly, William, and Ross Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *Quarterly Journal of Economics* 112(4): 1203–50.

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

Epstein, Larry G., and Martin Schneider. 2002. "Learning under Ambiguity." University of Rochester.

Epstein, Larry G., and Tan Wang. 1994. "Intertemporal Asset Pricing Under Knightian Uncertainty." *Econometrica* 62(2): 283–322.

Fernández, Carmen, Eduardo Ley, and Mark F. J. 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: Edward Arnold.

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

Giannoni, Marc P. 2002. "Does Model Uncertainty Justify Caution? Robust Optimal Monetary Policy in a Forward-Looking Model." *Macroeconomic Dynamics* 6(1): 111–44.

Gilboa, Itzhak, and David Schmeidler. 1989. "Maximin Expected Utility with Non-Unique Priors." *Journal of Mathematical Economics* 18(2): 141–53.

Hamilton, James D. 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle." *Econometrica* 57(2): 357–84.

Hansen, Lars Peter, and Thomas J. Sargent. 2001a. "Acknowledging Misspecification in Macroeconomic Theory." *Review of Economic Dynamics* 4(3): 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.' " Hoover Institution, Stanford University.

———. 2002c. "Robust Control and Filtering of Forward-Looking Models." Hoover Institution, Stanford University.

Hansen, Lars Peter, Thomas J. Sargent, and Thomas Tallarini. 1999. "Robust Permanent Income and Pricing." *Review of Economic Studies* 66: 873–907.

Heckman, James J. 2001a. "Accounting for Heterogeneity, Diversity and General Equilibrium in Evaluating Social Programmes." *Economic Journal* 111(475): F654–99.

———. 2001b. "Micro Data, Heterogeneity, and the Evaluation of Public Policy: Nobel Lecture." *Journal of Political Economy* 109(4): 673–748.

Hoeting, Jennifer A., and others. 1999. "Bayesian Model Averaging: A Tutorial." *Statistical Science* 14(64): 382–401.

Hurwicz, Leonid. 1951. “Some Specification Problems and Applications to Econometric Models.” *Econometrica* 19(3): 343–44.

Kadane, Joseph B., Ngai Hang Chan, and Lara J. Wolfson. 1996. “Priors for Unit Root Models.” *Journal of Econometrics* 75(1): 99–111.

Keynes, John Maynard. 1921. *A Treatise on Probability*. London: McMillan.

———. 1940. “On a Method of Statistical Business-Cycle Research: A Comment.” *Economic Journal* 50(197): 154–56.

King, Robert G., 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 Alternative Institutional Measures.” *Economics and Politics* 7(3): 207–27.

Knight, Frank H. 1921. *Risk, Uncertainty and Profit.* New York: Houghton Mifflin.

Knox, Thomas A. 2002. “Learning How to Invest When Returns Are Uncertain.” Harvard Business School.

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

———. 1983. “Let’s Take the Con Out of Econometrics.” *American Economic Review* 73(1): 31–43.

Leamer, Edward E., and Herman Leonard. 1983. “Reporting the Fragility of Regression Estimates.” *Review of Economics and Statistics* 65(2): 306–17.

Leeper, Eric M., Christopher A. Sims, and Tau Zha. 1996. “What Does Monetary Policy Do?” *BPEA*, *2:1996,* 1–78.

Levin, Andrew T., and John C. Williams. Forthcoming. “Robust Monetary Policy with Competing Reference Models.” *Journal of Monetary Economics*.

Levine, Ross, and David Renelt. 1992. “A Sensitivity Analysis of Cross-Country Growth Regressions.” *American Economic Review* 82(4): 942–63.

Lindley, Dennis V. 1990. “The 1988 Wald Memorial Lectures: The Present Position in Bayesian Statistics.” *Statistical Science* 5(1): 44–89.

Madigan, David, and Adrian E. Raftery. 1994. “Model Selection and Accounting for Model Uncertainty in Graphical Models Using Occam’s Window.” *Journal of the American Statistical Association* 89(428): 1535–46.

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

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

Manski, Charles F. 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.

———. Forthcoming. "Social Learning from Private Experiences: The Dynamics of the Selection Problem." *Review of Economic Studies*.

Marimon, Ramon. 1997. "Learning from Learning in Economics." In *Advances in Economics and Econometrics: Theory and Applications*, edited by David M. Kreps and Kenneth F. Wallis. Cambridge University Press.

Maritz, J. S., and T. Lwin. 1989. *Empirical Bayes Methods*, 2d ed. London: Chapman and Hall.

McAleer, Michael, Adrian R. Pagan, and Paul A. Volker. 1985. "What Will Take the Con Out of Econometrics?" *American Economic Review* 75(3): 293–307.

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

McConnell, Margaret M., and Gabriel Perez-Quiros. 2000. "Output Fluctuations in the United States: What Has Changed Since the Early 1980's?" *American Economic Review* 90(5): 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(1): 85–110.

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

Otrok, Cristopher. 2001. "On Measuring the Welfare Cost of Business Cycles." *Journal of Monetary Economics* 47: 61–92.

Phillips, Peter C. B. 1991. "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends." *Journal of Applied Econometrics* 6(4): 333–64.

Raftery, Adrian E. 1995. "Bayesian Model Selection in Social Research." *Sociological Methodology* 25: 111–95.

Raftery, Adrian E., David Madigan, and Jennifer A. 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 B. 1980. "Using Empirical Bayes Techniques in the Law School Validity Studies." *Journal of the American Statistical Association* 75(372): 801–16.

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

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

Sargent, Thomas J. 1999. *The Conquest of American Inflation.* Princeton University Press.

Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: Dover Publications.

Schmitt-Grohé, Stephanie, and Martin Uribe. 2000. "Optimal Fiscal and Monetary Policy under Imperfect Competition." Rutgers University.

———. 2002. "Anticipated Ramsey Reforms and the Uniform Taxation Principle: The Role of International Financial Markets." Rutgers University.

———. Forthcoming. "Optimal Fiscal and Monetary Policy under Sticky Prices." *Journal of Economic Theory*.

Sims, Christopher A. 1988. "Bayesian Skepticism on Unit Root Econometrics." *Journal of Economic Dynamics and Control* 12(213): 463–74.

———. 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, 2:2002,* 1–40.

Slovic, Paul, and Amos Tversky. 1974. "Who Accepts Savage's Axioms?" *Behavioral Science* 19(6): 368–73.

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

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

Taylor, John B. 1996. "How Should Monetary Policy Respond to Shocks While Maintaining Long-Run Price Stability?" In *Achieving Price Stability.* Kansas City, Missouri: 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(3): 655–79.

———. 1999b. *Monetary Policy Rules.* University of Chicago Press.

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

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

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

Wasserman, Larry. 2000. "Bayesian Model Selection and Model Averaging." *Journal of Mathematical Psychology* 44(1): 92–107.

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