

NBER WORKING PAPER SERIES

MODEL UNCERTAINTY AND POLICY EVALUATION:
SOME THEORY AND EMPIRICS

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

Working Paper 10916
<http://www.nber.org/papers/w10916>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2004

We thank the John D. and Catherine T. MacArthur Foundation and National Science Foundation, Vilas Trust and University of Wisconsin Graduate School for financial support. We are especially grateful to Ethan Cohen-Cole, Giacomo Rondina and Chih Ming Tan for outstanding research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

© 2004 by William A. Brock, Steven N. Durlauf, and Kenneth D. West. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Model Uncertainty and Policy Evaluation: Some Theory and Empirics
William A. Brock, Steven N. Durlauf, and Kenneth D. West
NBER Working Paper No. 10916
November 2004
JEL No. C5, E5

ABSTRACT

This paper explores ways to integrate model uncertainty into policy evaluation. We first describe a general framework for the incorporation of model uncertainty into standard econometric calculations. This framework employs Bayesian model averaging methods that have begun to appear in a range of economic studies. Second, we illustrate these general ideas in the context of assessment of simple monetary policy rules for some standard New Keynesian specifications. The specifications vary in their treatment of expectations as well as in the dynamics of output and inflation. We conclude that the Taylor rule has good robustness properties, but may reasonably be challenged in overall quality with respect to stabilization by alternative simple rules that also condition on lagged interest rates, even though these rules employ parameters that are set without accounting for model uncertainty.

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

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

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

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

Friedrich von Hayek¹

I. Introduction

This paper explores issues related to the analysis of government policies in the presence of model uncertainty. Within macroeconomics, increasing attention is being given to the positive and normative implications of model uncertainty. One major direction of this work has been initiated by the seminal contributions of Hansen and Sargent (2001a,b,2002,2003) on robustness in policy analysis. Examples of contributions to this research program include Giannoni (2002), Marcellino and Salmon (2002), Onatski and Stock (2002) and Tetlow and von zur Muehlen (2001) and our own Brock and Durlauf (2004a,b) and Brock, Durlauf and West (2003).² In this approach, model uncertainty is defined relative to a given baseline model; specifically, a space of possible models is constructed by considering all models that lie within some distance ε of the baseline. In evaluating policies, the loss associated with a given policy is determined relative to the least favorable model in the model space, i.e. preferences are assumed to follow a minimax rule with respect to model uncertainty. As such, this program follows the approach to decisionmaking initiated by Wald (1950).

¹von Hayek (1942, p.290).

²A number of the ideas in this literature originally appear in an unpublished working paper by Peter von zur Muehlen, reprinted in von zur Muehlen (2001).

Our approach to model uncertainty analyzes model spaces that are non-local in the sense that we do not require that the different models are close to each other according to some metric. For many macroeconomic contexts, it seems clear that model uncertainty is sufficiently severe that very disparate models should be regarded as potential candidates for the true or best model. In the context of monetary policy, there has been no resolution of the role of expectations in determining the effects of policies on macroeconomic outcomes; some authors favor backward looking models which eschew any role for expectations (e.g. Rudebusch and Svensson (1999)) while some prefer forward looking models, (e.g. Woodford (2003)) and some advocate hybrid models with both forward and backwards looking features (e.g. Galí and Gertler (1999)). Model uncertainty also exists within these classes. For the classes of models that employ expectations, one finds differences with respect to how they are formulated, with disagreement about the use of rational expectations versus survey-based measures, for example. Yet another source of differences concerns the dynamic specification of a model in terms of lag length structure.

Formally, we treat uncertainty with respect to the true model in a fashion that is symmetric to other forms of uncertainty. From this perspective, the analysis of policies based upon a single model may be thought of as producing conditional probability statements in which one of the conditioning elements is the model. This approach to understanding how models effect policy evaluation leads to the use of model averaging methods, in which one first evaluates the conditional probability of some unknown object of interest given data and a choice of model and second eliminates this conditioning on a model by integrating out the model “variable.” Eliminating this dependence amounts to taking weighted averages of the model-specific probabilities, where the weights correspond to the probabilities of each model being the correct one. Model averaging represents an important recent development in the statistics literature; major contributions include Draper (1995) and Raftery, Madigan and Hoeting (1997). Model averaging methods require the specification of probabilities across models in order to compute posterior probabilities concerning parameters or other unknowns (such as forecasts) of interest.

Within the economics literature, these model averaging methods are achieving increasing prominence. Areas of application include economic growth (Brock and Durlauf (2001), Brock, Durlauf and West (2003), Doppelhoffer, Miller and Sala-i-Martin (2000) and Fernandez, Ley, and Steel (2001)), finance (Avramov (2002)), and forecasting (Garratt, Lee, Pesaran and Shin (2003) and Wright (2003a,b)) Some initial work on applications to monetary policy evaluation appears in Brock, Durlauf and West (2003).

While model averaging is a powerful tool in addressing model uncertainty, one can imagine contexts in which a policymaker will want more information than simply a summary statistic of the effects of a policy on outcomes where model dependence has been integrated out. For example, a policymaker may be interested in policies whose effects are relatively insensitive to which model is the correct one. Alternatively, a policymaker may wish to engage in model selection, and would like to know how this selection affects the likely efficacy of the policy. One reason for this is that a policymaker may not wish to adjust policies in response to the updating of model probabilities. For this reason, we believe that proper reporting of the effects of model uncertainty should also include descriptions of how model choice affects the form of a policy rule and the payoffs associated with that rule. This dependence leads us to calculate statistics that measure the degree of *outcome dispersion*, which characterizes how the losses associated with a model-specific optimal policy rule depends on the model, and *action dispersion*, which measures how the optimal policy differs across alternative models in a model space.

Our work is most closely related to Levin and Williams (2003). This analysis compares policy rules under theoretically distinct models; models are averaged by assigning equal weights to each model; this approach differs from Bayesian averaging as their weights do not represent posterior model probabilities. Nevertheless, this paper is important as it is the first extensive analysis of model averaging methods as applied to monetary policy. A significant virtue of Levin and Williams over our paper is that they are able to work with a much richer theory set than we do, in particular they include a model-consistent forward-looking model in their model space. On the other hand, they do not address the implications of model uncertainty that arise because of dynamics. Our

work is complementary to Cogley and Sargent (2004) who consider US monetary policy but with a positive rather than a normative focus. Cogley and Sargent consider adjustments to US monetary policy generated by changes in the weights assigned by the Federal Reserve to different models of inflation, showing that such model uncertainty helps explain the dynamics of inflation after 1970.

Our major concern in the empirical work in this paper is with the appropriate way to present the results of policy evaluation exercises. One obvious way to think about this problem is simply to compute expected losses under different policies where the expectation calculations account for model uncertainty. This approach requires the specification of prior probabilities on the space of possible models. Alternatively, one can apply a minimax criterion even though the model space we study is non-local. As argued in Brock, Durlauf, and West (2003), one may interpret Leamer's (1983) extreme bounds analysis as doing this. However, our purpose is not to defend a particular way in which decisions respond to model uncertainty but rather to describe ways to report predictions concerning policy effects in a way that communicates how model uncertainty affects these predictions. We will therefore emphasize some informal quantitative and visual tools to communicate how model uncertainty enters policy evaluation.

We apply our general analysis to some standard questions on monetary economics. In our empirical analysis, we consider two classes of standard New Keynesian models. Models in each class include three equations: a dynamic IS curve relating output to a real interest rate; a dynamic Phillips curve relating inflation to output and expected inflation; and a monetary policy (Taylor) rule relating the interest rate to output, inflation and a lagged interest rate. The two classes differ in their treatment of expectations. Our backwards class, which builds on Rudebusch and Svensson (1999), treats expected inflation as a distributed lag on past inflation. Our hybrid class, which builds on Rudebusch (2002), uses survey data on expected inflation in estimation but assumes model-consistent expectations in evaluation of alternative monetary policies. Within a given class, models vary only in terms of the number of lags included on the right hand sides of the IS and Phillips curves. We consider the effects of alternative monetary policy rules using a loss function based on a weighted average of variances of

output, inflation and interest rates; we refer to the losses incurred under this specification as risk. Our analysis of the model space reveals that the hybrid models possess a posterior probability that is an order of magnitude higher than that of the backwards looking models. So while we do average within classes, we do not average across the model classes, and rather report results for each class separately. We do so because our model classes are defined around models that themselves were data mined for distinct model spaces. We regard the question of how to construct model spaces around data mined models to be an important outstanding research question.

We conduct three different empirical analyses. First, we consider the behavior of the losses associated with the classic Taylor (1993) rule when model uncertainty is present. Our findings suggest that risk estimates for the Taylor rule are quite robust in the sense that our risk estimates show relatively little variation across models. Second, we compare the performance of the Taylor rule with the performance of an interest rate rule that sets current rates as a function of the lagged rate, current inflation, and current output. We choose the parameters of the rule such that the parameters are optimal for the model with the maximum posterior probability in each of our classes. We find that for the backwards models, the optimized rule systematically dominates the Taylor rule, except for a small (in posterior probability sense) set of models where the optimized rule induces instability in the system. For the hybrid class, the optimized rule uniformly dominates the Taylor rule. Our final exercise considers how optimal three-variable interest rate rules vary across models. In this exercise, we compute optimal rules and associated risks for each model in the two model classes. Our analysis of outcome and action dispersion is largely visual as it consists of the presentation of dispersion figures. As such, it is somewhat hard to identify simple messages from the exercise. One conclusion we do draw is that there appears to be some systematic relationship between the coefficients in the model-specific optimal rules and model complexity.

The paper is organized as follows. Section II of this paper describes our basic framework. Section III contains our various empirical exercises. Section IV provides some interpretation of the findings in the context of a general dynamic linear model. Section V provides conclusions.

II. Incorporating model uncertainty into statistical analyses

Our basic argument concerning the analysis of policy in the presence of model uncertainty is that such uncertainty should be explicitly incorporated in the calculation of the effects of a policy. In other words, we argue that from a decision-theoretic perspective, model uncertainty is not a property that should, via model selection exercises, be resolved prior to the evaluation of a policy rule, but rather is a component of that evaluation. To see why this is so, we follow the discussion in Brock, Durlauf, and West (2003); other analyses that advocate an explicit decision-theoretic approach to the analysis of data in economics include Chamberlain (2001) and Sims (2002). This analysis is a straightforward application of standard statistical decision theory arguments, cf. Berger (1987).

i. general framework

Suppose that a policymaker wishes to evaluate the effect of a policy rule p on an outcome θ . We assume that the policymaker's assessment of the outcome depends only on the outcome so that one can separate the preferences of the policymaker from the probability measure characterizing θ given the policy. In assessing policies, the question of model uncertainty arises in the context of specifying the information set on which the assessment is conditioned. Typically, one begins with a specification of the data generating process, i.e.

$$\theta = m(p, \beta_m, \eta) \tag{1}$$

where m denotes a model, p is a policy, β_m is a vector of parameters that indexes the model and η is a set of unobservable shocks that affect θ . It may be assumed, without loss of generality, that when evaluating policies, the data generating process and

probability measure for the innovation, μ_η are known even though the realizations of the shocks are not, so that policies are evaluated based on the conditional probability measure

$$\mu(\theta|p, m, \beta_m) \quad (2)$$

This formulation indicates the first level at which the effects of policies are uncertain. Even if the data generating process and associated parameters are known, there is uncertainty due to the unobservability of η .

Eq. (2) implies that a policymaker possesses a great deal of information about the data generating process. Such information is typically not available to the researcher, and its absence must be accounted for to provide appropriate statements about the effects of a policy. The relaxation of the information implicitly assumed in (2) may be done in two steps. First, assuming that the model is known, there is typically uncertainty about the values of the model parameters β_m . Operationally, this means that one computes

$$\mu(\theta|p, m, d) \quad (3)$$

The difference between (2) and (3) is that in (3) one is implicitly using the available data d to construct estimates of the model parameters. For macroeconomic problems, this is generally regarded as a second-order; exceptions to this include Giannoni (2001) and Onatski and Williams (2003). While we do not address parameter uncertainty in our empirical examples, we note that the lack of importance of parameter uncertainty has by no means been established as an empirical matter and is in fact contradicted by Gianonni's and Onatski and Williams's findings; this is a topic that warrants further research.

For our purposes, the key issue of interest is how to move beyond (3) to account for uncertainty in the specification of the data generating process, which we will refer to as model uncertainty. This level of uncertainty captures that of strong knowledge concerning economic theories, functional form specification (including threshold effects, switching regimes, etc.) and heterogeneity in the data generating processes for individual

observations. Brock, Durlauf and West (2003) provide a typology of forms of model uncertainty along these lines. One goal of a policy evaluation may be the calculation of

$$\mu(\theta|p, d). \tag{4}$$

In other words, one way a policymaker can deal with model uncertainty is to treat it as another type of unobservable similar to η and β_m and evaluate policies in a way that accounts for this.

As recognized originally in Leamer (1978) and developed in subsequent work such as Draper (1995), this idea may be operationalized using standard probability arguments to eliminate the conditioning on m that is present in (3). To do this, suppose that an analyst is working with a space M of possible data generating processes. We will implicitly assume that the true model is an element of this space when discussing how we interpret empirical findings; none of the empirical findings we present are themselves dependent on that assumption.³ Without loss of generality, we take the space to be countable.

Standard application of conditional probability arguments implies that the $\mu(\theta|p, d)$ may be characterized as follows:

$$\mu(\theta|p, d) = \sum_m \mu(\theta|p, m, d) \mu(m|d) \tag{5}$$

In this expression, $\mu(m|d)$ is known as the posterior probability of model m given data d . From the perspective of (5), model uncertainty is treated in a fashion that is symmetric to any other source of uncertainty in θ .

Eq. (5) reveals how the incorporation of model uncertainty into policy analysis requires the calculation of a class of objects, posterior model probabilities, which simply do not appear when one evaluates policies after engaging in model selection. To

understand what these probabilities mean, by Bayes' rule, these probabilities are the product of two terms, i.e.

$$\mu(m|d) \propto \mu(d|m)\mu(m) \tag{6}$$

where $\mu(d|m)$ is the likelihood of the data given model m and $\mu(m)$ is a prior probability assigned to model m . This derivation illustrates two features concerning the role of model uncertainty in policy evaluation.

First, if one starts with a space of possible models which is constructed without knowledge of which models fit particularly well, then model averaging can ameliorate problems associated with data mining. Eq. (5) indicates how probability calculations can employ all models in the model space, incorporating the relatively greater likelihood of some models versus others via the $\mu(d|m)$ terms. Hence, the standard problem of data mining, drawing inferences about a model without accounting for its selection, does not arise. This observation requires two caveats. First, it is important in constructing the $\mu(d|m)$ terms to avoid overweighting more complex models simply because of their superior goodness of fit. As we shall see below, model complexity penalties (in our case, based on the BIC) are needed when calculating posterior model probabilities. Second, in some cases it may not be possible or practical to analyze the set of all possible models. Hence, data mining problems may occur because of limits in the analysis that exist in the model space in the way we have described.

Second, the issue of model selection does not arise when one takes the averaging perspective. Heuristically, one may understand model selection exercises as choosing a model based on its relative goodness of fit (adjusted for model complexity). In the context of our approach, model selection of this type is equivalent to placing a posterior probability of 1 on the model with the highest posterior probability. Thought of this way, it is easy to see why model selection can lead to very misleading assessments of policy

³Bernardo and Smith (1994) discuss the interpretation of model spaces under alternative assumptions as to whether the true model is in the space.

efficacy. For example, model uncertainty calculations avoid situations where one model may far outstrip others by a selection criterion, yet the posterior model probability is small relative to the space as a whole.

Third, any analysis of model uncertainty will be dependent on a researcher's prior beliefs about the relative plausibility of different models, as quantified through the prior probabilities $\mu(m)$. Very little work has been done on the question of appropriately formulating priors over model spaces. Most papers assign a uniform prior across the model space. One alternative, suggested by Doppelhofer, Miller, and Sala-i-Martin (2000) penalizes complex models by assigning relatively lower prior weights to them. Brock, Durlauf, and West (2003) discuss ways to use economic information to structure priors that reflect theoretical, specification, and parameter heterogeneity differences between models. However, this is a question that needs much more research.

Calculations of this type make clear how model uncertainty affects policy evaluation. Suppose that a policy maker evaluates policies according to a risk function⁴ $R(\theta)$ and that the policymaker evaluates a policy rule based on the expected loss it generates. Standard policy analyses calculate

$$E(R(\theta)|p, m, d) = \int_{\Theta} R(\theta) \mu(\theta|p, m, d) d\theta \quad (7)$$

whereas an analysis that allows for model uncertainty should calculate

$$E(R(\theta)|p, d) = \int_{\Theta} R(\theta) \mu(\theta|p, d) d\theta \quad (8)$$

In contexts such as stabilization policy, one usually is interested in the first two moments of $\mu(\theta|p, d)$. These moments were originally computed by Leamer (1978) and are discussed in great detail in Draper (1995):

⁴We refer to a risk function rather than a loss function in order to use language that is standard in the monetary policy literature.

$$E(\theta|p, d) = \sum_m \mu(m|d) E(\theta|p, m, d) \quad (9)$$

and

$$\begin{aligned} \text{var}(\theta|p, d) &= E(\theta^2|p, d) - (E(\theta|p, d))^2 = \\ &= \sum_{m \in M} \mu(m|d) \text{var}(\theta|p, m, d) + \sum_{m \in M} \mu(m|d) (E(\theta|p, m, d) - E(\theta|p, d))^2 \end{aligned} \quad (10)$$

ii. model uncertainty and ambiguity aversion

Our discussion has so far treated model uncertainty in a way that is equivalent to innovation uncertainty (i.e. lack of knowledge of η) and parameter uncertainty. There are reasons to believe that one may not want to assume this equivalence. One of these reasons derives from the body of experimental work that is associated with the Ellsberg paradox. Consider two scenarios: in scenario A, a bet may be placed on the color of a ball drawn from an urn when the distribution of colors is known to be 50/50 between black and red whereas in scenario B a bet may be placed on the color of a ball drawn from an urn where the distribution between black and red is not known, but where the subject can choose the color. The Ellsberg paradox refers to the observation that in various experiments, subjects place a higher value on the former bet; the paradox occurs since by symmetry, the fact that one can choose which color ball to bet on makes it impossible to differentiate the second bet from the first bet in an expected value sense. Such observations have led to a recent literature on ambiguity aversion, exemplified by Gilboa and Schmeidler (1989) and Epstein and Wang (1994). Following Epstein and Wang (1994), ambiguity aversion can be introduced by considering the modified expected loss function

$$(1-e) \int_{\Theta} l(\theta) \mu(\theta|d) d\theta + e \left(\sup_{m \in M} \int_{\Theta} l(\theta) \mu(\theta|m, d) d\theta \right) \quad (11)$$

This loss function places an additional weight on the least favorable model in the model space beyond that which is done in a standard expected loss calculation.⁵ This function nests the expected loss approach ($e=0$) and the minimax approach ($e=1$) that is employed in the macroeconomics robustness literature, cf. Hansen and Sargent (2001a,b,2002,2003).

iii. model uncertainty and stabilization policy

We now specialize these formulas for the analysis of stabilization policies. To do this, we consider the scalar case where the policymaker is interested in stabilizing output relative to trend, y_t . We assume that a policymaker evaluates rules according to their limiting effect y_∞ , specifically the policymaker's loss function is

$$\text{var}(y_\infty | p, d) \tag{12}$$

This loss function is timeless in the sense of Woodford (2003) and thus avoids problems of time inconsistency. We assume that the policy cannot affect the long-run mean of the series, so that

$$E(y_\infty | p, d) = 0 \quad \forall p \tag{13}$$

This is a substantive economic assumption and one that is frequently built into macroeconomics models, for example to reflect a long run Phillips curve. Under this assumption

$$\text{var}(y_\infty | p, d) = \sum_{m \in M} \mu(m|d) \text{var}(y_\infty | p, m, d) \tag{14}$$

⁵A remarkable early formulation of this type appears in Hurwicz (1951).

Relative to (10), the second term on the right hand side (RHS) disappears when (13) holds.

In the context of analyzing stabilization policies, one can further observe that the overall variance associated with a given policy, $Var(y_\infty | p, d)$, may be contrasted with two other calculations which are suggested by our discussion:

$Var(y_\infty | p, m, \beta_m)$ = overall within-model variance due to unobserved innovations; this level of variance is irreducible in the sense that it is present even if a model and associated parameters are known

$Var(y_\infty | p, m, d)$ = overall within-model variance due to parameter uncertainty given a model

As one moves from uncertainty due to innovations and parameters to uncertainty that also reflects lack of knowledge of the true model, one moves from conventional model exercises to the approach we advocate. Put differently, if one engages in model selection, one typically computes $Var(y_\infty | p, m, \beta_m)$ or $Var(y_\infty | p, m, d)$ whereas we would argue the correct object for study in policy analysis is $Var(y_\infty | p, d)$.

Finally, we consider how to evaluate uncertainty about the variance we have described; we focus specifically on the “variance of the variance” associated with a given policy. While a mean/variance loss function is not affected by this calculation, other preferences structures are. To perform these second order variance calculations, notice that by (13), $Var(y_\infty | p, d) = E(y_\infty^2 | p, d) - [E(y_\infty | p, d)]^2$. We can thus use formulas (9) and (10) to calculate $Var(y_\infty^2 | p, d)$. Since $E(y_\infty^2 | p, d)$ is, unlike $E(y_\infty | p, d)$, dependent on p , one has

$$\begin{aligned} & \text{var}(y_\infty^2 | p, d) = \\ & \sum_{m \in M} \mu(m|d) \text{var}(y_\infty^2 | p, m, d) + \sum_{m \in M} \mu(m|d) \left(E(y_\infty^2 | p, m, d) - E(y_\infty^2 | p, d) \right)^2 \end{aligned} \quad (15)$$

The second term on the RHS of (15) captures the distinct role that model uncertainty plays in assessing the payoff associated with a policy. The first term represents the uncertainty contribution given the models. This decomposition into a within-model and across-model uncertainty corresponds to the analysis in Gustafson and Clarke (2004). Notice that the only reason why $E(y_\infty^2 | p, d, m) - E(y_\infty^2 | p, d)$ is nonzero is variability across models.

These calculations lead to a hierarchical view of policy assessment. As we have claimed above, conventional policy evaluation exercises calculate either $\text{Var}(y_\infty | p, m, \beta_m)$ or $\text{Var}(y_\infty | p, m, d)$ where the model m is chosen by some criterion that trades goodness of fit against model complexity. Such calculations are of course important. What we argue is that in addition to such calculations, one should also compute $\text{Var}(y_\infty | p, d)$, which describes the consequences of the same policy without the assumption that the model selection exercise has identified the correct model. The discrepancy between these two measures will provide a metric for the economic significance of model uncertainty. Notice that there is no necessary ordering between $\text{Var}(y_\infty | p, d, m, \beta_m)$, $\text{Var}(y_\infty | p, d, m)$ and $\text{Var}(y_\infty | p, d)$ as model uncertainty may lower the variance associated with a policy if the policy works better for those parameters that have not been assumed or for a different model. It is possible for the introduction of model uncertainty to reverse the relative rankings of models.

iv. implementation issues

a. priors and the reporting of results

The calculations we have described require the specification of prior probabilities for the elements of the model space M . The construction of priors continues to be a knotty problem in Bayesian statistics. One difficulty in the construction of priors derives from the difficulties inherent in translating vague prior beliefs possessed by a researcher into probabilities. This difficulty has led to a large literature on Bayesian probability elicitation, an approach that has not been pursued in the model uncertainty context. Most studies of model uncertainty and model averaging assume that all elements in M possess equal prior probabilities, a standard assumption when one wants to employ a noninformative prior, i.e one that expresses ignorance.⁶ Other authors have modified the equal probability assumption either by assuming the model probabilities are themselves random, which in essence makes the prior a mixture distribution (Brown, Vannucci, and Fearn (1998)) or by assigning higher prior probabilities to simpler models (Doppelhofer, Miller, and Sala-i-Martin (2000)). None of these approaches use social science reasoning to construct priors. Brock, Durlauf, and West (2003) argue that priors should possess a hierarchical structure that reflects the differences between theory uncertainty and specification uncertainty conditional on a theory. It is unclear that these different approaches are of major importance operationally.

An alternative perspective is that the goal of a policy evaluation analysis is to communicate to a policymaker the effects of a policy under alternative assumptions rather than to perform expected loss calculations per se. As we have argued, assumptions about the theoretical basis and specification of the model of the phenomenon of interest are of primary importance in this respect. To the extent this is true, and recognizing the possibility that ambiguity aversion means that a policymaker may react to model uncertainty differently from parameter uncertainty, for example, then the averaging approach may not be sufficiently informative. A policymaker may want to know if there are outlier models in the sense that a policy works particularly poorly when they are correct. Notice that this is not the same thing as asking whether certain models are outliers in terms of certain parameter values, overall goodness of fit, etc. For this reason,

⁶There are many conceptual problems in defining what it means for a prior to be uninformative; these issues are not germane to our discussion.

we argue that a significant part of a policy evaluation exercise is the presentation of different perspectives on how model uncertainty affects one's conclusions. We are therefore concerned to identify useful statistics and visual representations of policy effects as they vary across models.

b. Bayesian versus frequentist

Our discussion has been explicitly Bayesian in that our analysis has focused on the construction of probability measures on the unknowns θ given observed data and prior information, i.e. $\mu(\theta|d)$. These calculations, in turn, employed Bayesian within-model posterior probabilities $\mu(\theta|d,m)$. That being said, the logic of our model averaging arguments really only depend on the use of posterior model probabilities $\mu(m|d)$. If one can identify an interpretable way of constructing these model probabilities, then one can integrate these with frequentist objects in order to address model uncertainty without fully committing to Bayesian methods. For example, if one is interested in constructing an estimate $\hat{\theta}$ which is not model-dependent, this can be done via

$$\hat{\theta} = \sum_{m \in M} \hat{\theta}_m \mu(m|d) \quad (16)$$

Doppelhofer, Miller, and Sala-i-Martin (2000), who perform such calculations in the context of OLS regression parameters when there is uncertainty about the choice of controls, call this approach Bayesian averaging of classical estimates (BACE); Brock, Durlauf, and West (2003) refer the general approach of averaging frequentist objects using model weights as a pseudo-frequentist procedure. What is important, of course, is not the terminology, but the idea that incorporation of model uncertainty into a data exercise can provide interpretable results. This is extremely important since frequentist

methods dominate policy evaluation analysis. We employ this pseudo-frequentist approach in the empirical section of this paper.

vi. beyond model averaging: outcome dispersion and action dispersion

As suggested in the Introduction, in evaluating the role of model uncertainty in policy assessment, we believe it is useful to think about two concepts: outcome dispersion and action dispersion. *Outcome dispersion* captures the variation in loss that occurs when one considers different models. When working with a fixed policy, the variation of losses under the policy traces out the range of the loss function, where the latter is interpreted as a function of the policy. Averaging calculations can thus be treated as data reductions of the support of the loss function; a data reduction in which a (posterior probability) weighted sum of the range is computed.

When a policy is allowed to depend on a model, one can define an analogous notion of *action dispersion*. Each model induces a distinct policy, so the model space traces out a range of policies. For example, one can compute how the parameters of a simple monetary policy rule, say one that maps last period's Federal Funds rate, the current inflation level and the current output level into this period's Federal Funds rate, varies across models. Why might such information be of use to a policymaker? One reason is that calculations of action dispersion can reveal how sensitive a policy rule is to model choice. To the extent that a policymaker decides to condition policies on a model, action dispersion can reveal the extent to which this matters. In turn, one can argue that a desideratum of a policy rule is that its formulation is relatively insensitive to certain details of the economic environment in which it is applied. Giannoni and Woodford (2002) make this idea precise in a theoretical context; our calculations of action dispersion provide an empirical representation to their ideas.

III. Model uncertainty and assessment of simple monetary policy rules

In this section, we provide an illustration of the methodological discussion using a simple empirical example; the example extends work in Brock, Durlauf, and West (2003).

i. framework

We suppose that a monetary policymaker is contemplating the choice of parameters in a simple monetary policy rule. This class of rule is studied in many papers, a thorough example is Levin, Wieland and Williams (1998). Denoting the output gap as y_t , inflation as π_t and the nominal interest rate on 1-period government bonds as i_t , we assume that the policymaker employs a nominal interest rate rule

$$i_t = g_\pi \pi_t + g_y y_t + g_i i_{t-1} \quad (17)$$

Following standard assumptions and terminology in the monetary rules literature, losses are calculated via a risk function R , defined as

$$R = var(\pi_\infty) + \lambda_y var(y_\infty) + \lambda_i var(\Delta i_\infty) \quad (18)$$

In our risk calculations, we will always assume $\lambda_y = 1.0$ and $\lambda_i = 0.1$. This choice of weights is arbitrary but is in the range assumed by earlier literature using similar loss functions, e.g. Levin and Williams (2003).

Our alternative models represent examples of the New Keynesian model expounded in Woodford (2003). The particular representations we employ are taken from Rudebusch and Svensson (1999) and Rudebusch (2002). These models may be understood as two equation systems. The first component of the system is an IS curve that relates output to real interest rates and an unobservable disturbance $u_{IS,t}$:

$$y_t = \alpha_{y1} y_{t-1} + \alpha_r (\bar{i}_{t-1} - E_{t-1} \bar{\pi}_{t+3}) + \left[\sum_{j=2}^4 \alpha_{yj} y_{t-j} \right] + u_{IS,t} \quad (19)$$

where $\bar{i}_t = \frac{1}{4} \sum_{j=0}^3 i_{t-j}$, $\bar{\pi}_t = \frac{1}{4} \sum_{j=0}^3 \pi_{t-j}$ and $u_{IS,t}$ is an unobservable disturbance. The second component is a Phillips curve that relates inflation to expected inflation, lagged inflation, lagged output and an unobservable disturbance $u_{PC,t}$. The weights on inflation are constrained to sum to unity in order to ensure that the curve is vertical in the long run.

$$\pi_t = \beta_0 E_{t-1} \bar{\pi}_{t+3} + (1 - \beta_0) \beta_{\pi_1} \pi_{t-1} + \beta_{y_1} y_{t-1} + \left[(1 - \beta_0) \sum_{j=2}^4 \beta_{\pi_j} \pi_{t-j} + \sum_{j=2}^4 \beta_{y_j} y_{t-j} \right] + u_{PC,t} \quad (20)$$

subject to $\beta_0 + (1 - \beta_0) \sum_{j=1}^4 \beta_{\pi_j} = 1$

In eq. (20) and throughout, we suppress inessential constants for expositional simplicity; these were included in all our empirical work.

Model uncertainty exists at two levels in our framework. The first level corresponds to our notion of theory uncertainty as it relates to the way in which expectations are formed by agents. First, backwards looking and hybrid models are differentiated by treatment of $E_{t-1} \bar{\pi}_{t+3}$. For backwards looking models,

$$E_{t-1} \bar{\pi}_{t+3} = .25 (\pi_{t-1} + \pi_{t-2} + \pi_{t-3} + \pi_{t-4}) \quad (21)$$

whereas for hybrid models.

$$E_{t-1} \bar{\pi}_{t+3} = \text{survey data on 1-year ahead inflation} \quad (22)$$

The backwards looking modeling follows Rudebusch and Svensson (1999) whereas the hybrid modeling follows Rudebusch (2002). As well, the backwards model sets the coefficient on expected inflation in the Phillips curve to 0 (i.e. $\beta_0 = 0$). We refer to the backwards and hybrid cases as our two classes of models.

Second, there is specification uncertainty that exists once one has conditioned a given theory. This uncertainty is modeled with respect to the terms in brackets in equations (19) and (20). Different lag structures correspond to alternative ways of capturing output and inflation dynamics; these dynamics are not constrained by economic theory but rather are included in order to capture serial correlation in the model errors. In each class of models, we estimate 4 different IS curves, with one, two, three and four lags of output on the RHS. We estimate 16 different Phillips curves, with one to four lags of output and one to four lags of inflation in the RHS. Thus within each class of models there are $64 = 4 \times 16$ specifications; each specification corresponds to a specific set of lag structures for the IS/PC system.

Under the assumption that policy is deterministic, we use estimated values for the parameters of the IS and Phillips curves to solve the model and compute values of the loss function under alternative policy parameters. Our analysis assumes that the IS and Phillips curves are structurally stable over the 1970-2002 sample. We are aware of evidence to the contrary, but leave this complication to future work. We also do not allow for one class of models to represent a better approximation of the underlying data generating process in some periods but not others.⁷ Our simplifications are made to facilitate the exposition of how one might incorporate model uncertainty in evaluating the losses associated with alternative policies. For each model and a given set of policy preference parameters λ_y and λ_π , we use a grid search procedure to solve for the values of g_π , g_y , and g_i that minimize the loss function (18).

We calculate losses as follows. For a given model m , let \hat{R}_m denote model risk, when uncertainty associated with estimated parameters is ignored. Let \hat{L}_m denote the BIC-adjusted likelihood for the model. For a given set of models, the expected risk \hat{R} associated with model uncertainty is

⁷See Brock and Hommes (1997) for a theoretical discussion of modeling epoch-dependent expectations formation in which individual agents make correlated investment decisions in information that collectively vary at different points in time and Pesaran, Pettenuzzo, and Timmermann (2004) for methods to identify different epochs.

$$\hat{R} = \sum_{m \in M} \hat{R}_m \mu(m|d) \quad (23)$$

We assume that all models within a model class have equal prior probability. While we would prefer to assign priors in ways that are suggested by economic reasoning, we have yet to develop a natural way to do so in this context. We also see no reason why more complicated models warrant smaller (or larger) priors than simpler ones. Our uniform prior assumption implies that $\mu(m|d)$ is proportional to \hat{L}_m so that

$$\hat{R} = \frac{\sum_{m \in M} \hat{R}_m \hat{L}_m}{\sum_{m \in M} \hat{L}_m} \quad (24)$$

We consider a number of ways to communicate the importance of model uncertainty in policy choice. In addition to various averaging calculations, we quantify our notions of outcome and action dispersion. Dispersion is measured in several ways, including support width (absolute value of the difference between the maximum and minimum values of the object of interest as it varies across models), standard deviation and interquartile range of risk across models. In reporting outcome dispersion, we acknowledge that one would like to consider outcome dispersion with respect to a range of policy preference structures but do not do so here. Finally, note that action dispersion is measured by dispersion in $\tilde{g}_\pi = \frac{g_\pi}{1-g_i}$ and $\tilde{g}_y = \frac{g_y}{1-g_i}$ and g_i . We employ the normalizations \tilde{g}_π and \tilde{g}_y in order to evaluate variation in the long-run effects of income and inflation on interest rates respectively.

As part of our goal is to report visual descriptions of the properties of the model space, we will associate each model with a number. This relationship is described in Appendix 1.

ii. data

All estimation is done using quarterly data from 1970:2 to 2002:4, with data from 1969:2 to 1970:1 used to provide lags. Apart from survey data, this is the same data studied in Brock, Durlauf, and West (2003). Inflation π_t is measured as the annualized change in the GDP deflator. The output gap y_t is computed as the difference between real GDP and the Congressional Budget Office's estimate of potential GDP. The interest rate i_t is the quarterly average Federal Funds rate. We constructed the survey expectations measure of $E_{t-1}\bar{\pi}_{t+3}$ from the median price expectations of the Survey of Professional Forecasters. Let $P_{t|t}^e$ denote the period t survey expectation of the GDP deflator (GNP deflator prior to 1992) in the current quarter and $P_{t+4|t}^e$ denote the expectation of the deflator four quarters (one year) from t . We set $E_{t-1}\bar{\pi}_{t+3} = 100 \times \log(P_{t+3|t-1}^e / P_{t-1|t-1}^e)$. For two quarters (1970:3 and 1974:3), $P_{t+4|t}^e$ was missing; we substituted an extrapolation of the three-quarter-ahead expectation $P_{t+3|t}^e$.

iii. basic properties of the model space

We first consider some properties of the model space. Table 1 presents regression results for the backward and hybrid specifications with the highest posterior probability. These are the models that would be selected if one were using the BIC criterion to choose one model within each class. The results are consistent with those for the backward specification of Rudebusch and Svensson (1999) and the hybrid specification of Rudebusch (2002). In the IS curve, the BIC-adjusted likelihood chooses three lags of output in the backwards specification, two lags in the hybrid specification. The sum of regression coefficients and the interest rate elasticity are similar in both specifications. In the Phillips curve, both specifications choose one lag of output. The backward specification uses three lags of inflation, while the hybrid combines the survey expectation with a single lag. (Recall that by construction, the sum of the lags (and lead,

for the hybrid specification) on inflation is 1.) The hybrid specification puts substantial weight on the survey expectation, with $\hat{\beta}_0 = 0.32$.

The maximum posterior probability hybrid model involves two fewer parameters than does this backwards model. For this reason, as well as some other quantitatively less important ones, the BIC-adjusted bivariate likelihood for the hybrid model is two orders of magnitude higher than that of the backward looking model (not reported in the table). We do *not* interpret the relative BIC-adjusted likelihoods as arguing for great posterior weight on hybrid versus backwards models. We came to this specification only after experimenting with various model consistent measures of expectations (not reported), and by choosing the very best fitting specification in Rudebusch (2002). For example, we do not include terms on forward looking output in the IS equation, because Rudebusch (2002) found these to not be significant. We return to this point below when we combine backward and hybrid models.

How do model probabilities differ across the model space? Table 2 presents summary statistics on the distribution of the posterior model probabilities across the 64 models in each of the two classes. To do this, we focus on the relative likelihoods of each model m within a class, defined as

$$P_m = \frac{\hat{L}_m}{\sum_{m \in C} \hat{L}_m} \quad (25)$$

where the sum in the denominator runs over the 64 models in a given class (backward or hybrid). By construction, $0 < P_m < 1$ and $\sum_{m \in C} P_m = 1$. In each class, the relative likelihood is clustered around a handful of models. Row (6) in Table 2 indicates that only 8 (backward) or 13 (hybrid) models have likelihood as much as 1/20 of the likelihood of the model with the highest posterior. We will designate this group of models as possessing “high” likelihoods or “high” posteriors in our subsequent discussion. The factor of 1/20 is made to facilitate highlighting those models most consistent with the data and follows ideas that have appeared elsewhere in the model averaging literature,

e.g. Gustafson and Clarke (2004); minor changes in the definition of what is meant by a posterior probability would not change any qualitative features of our discussion. Row 8 of Table 2 indicates that in each class of models the 16 models with highest posterior probability dominate the relative likelihood.

iv. the original Taylor rule revisited

Our first analysis using the model space considers the effects of model uncertainty on evaluation of the risk associated with the original Taylor (1993) rule⁸:

$$i_t = 1.5\pi_t + 0.5y_t \tag{26}$$

This rule may be evaluated with respect to outcome and action dispersion. Relative to our earlier discussion, action dispersion is by definition 0 since the rule is constant across specifications. Outcome dispersion is described in Table 3, which characterizes the way in which the risk associated with the original Taylor rule varies across the model space. Overall, for the class of backwards looking models, the risk values appear to be relatively stable. When one concentrates on relatively likely backwards models, the risk estimates are all in the range of 19.1 to 23.2; the same exercise for hybrid models yields the somewhat broader range of approximately 15.2 to 31.9. There do exist outlier models with very different risk values: the support for Taylor rule risk for the backwards models is appropriately 17.5 to 51.6 and the support for the hybrid models is 12.5 to 44.7. Row 8 of the Table provides the model averaging calculations, in which the model specific risk of the Taylor rule is averaged using posterior model probabilities according to (24). It is interesting to compare our model averaged risk estimates, 22.0 for the backwards class and 23.6 for the hybrid class, with the respective risks that occur for the maximum posterior probability models in each class, 23.2 and 24.1 respectively. The averaged numbers are lower, indicating that the Taylor rule works better for at least some models that would be ignored if one simply focused on the maximum posterior models.

⁸We report the demeaned version of the rule but used constants in the empirical work.

This exercise suggests that the Taylor rule generally has good outcome robustness properties.

v. comparing simple rules

As a second exercise, we consider the relative performance of the original Taylor rule against an optimized three-variable rule of the form (17). To do this, we calculate values of g_π , g_y and g_i which minimize the risk function (18) using the weights described below the equation for the backward model with the highest posterior probability and hybrid model with the highest posterior probability. As described in the next section, we found these parameters by a grid search. The results of the grid search are:

$$\text{backwards: } \tilde{g}_\pi = 3.2, \tilde{g}_y = 2.1, g_i = 0.2; \text{ hybrid: } \tilde{g}_\pi = 3.2, \tilde{g}_y = 4.7, g_i = 0.55 \quad (27)$$

Both policies are more aggressive than the original (1993) Taylor rule.

Our objective is to compare the performance of these rules with the Taylor rule. The optimized rules will of course outperform the Taylor rule when the posterior model with the highest probability in a model space is the true one; what we wish to ascertain is how this comparison is affected when one accounts for the presence of model uncertainty. In order to do these comparisons, we perform two sets of exercises. First, we compare the Taylor rule to the model-specific three-variable rule where the rule is computed for the same class on which the comparison is done. These comparisons mean that the policymaker is confident that his given choice of model class is the correct one, and is concerned only with misspecification within that class. Second, we do the same comparisons when the policymaker has chosen the wrong class. This means we compare the Taylor and three-variable rule optimized for the higher posterior backwards model on the class of hybrid models and vice versa. This exercise will be of interest to a policymaker who has tentatively chosen a model class but wishes to understand the costs if the other class in fact better captures salient features of the economy.

Figure 1 presents a graph of the relative risks of the optimized 3-variable and Taylor rules across the model space for both our exercises. Models are reported using the numbering described in Appendix 1. All relative risks are the ratios of the risk using the optimized rule to the risk using the Taylor (1993) rule, eq. (26). We depict those cases where the hybrid rule produced instability, which happened to occur for some backwards but no hybrid models, with a solid line truncated at 1.8; this is done for readability.

Figure 1 yields several interesting findings. We first focus on the two graphs in the first row, in which the policymaker is confident a given class of models is the correct one. As the Figure indicates, for the hybrid models, the optimized rule uniformly dominates the Taylor rule across the model space. Second, for the backwards models, the relative risk of the optimized 3-variable rule is either 40% smaller than the Taylor rule or greater than 1. An examination of the specific models for which the Taylor rule outperforms the optimized rule explains why this is happening. For this subset of models, the optimized rule produces instability in at least one of the state variables, thereby producing infinite risk. The possibility that a rule that is optimal for one model produces instability in another is, as a theoretical matter, not surprising, and has been recognized by other authors, cf. Levin and Williams (2003, pg. 953). How serious a problem is this? The posterior probability for the set of models for which the optimized rule produces instability is .003.⁹ Hence, the probability of instability appears to be quite small. Because of the loss function that is assumed, if any of these models receives a positive weight in the expected loss calculation, the case that the Taylor rule will be preferred. This is an example where we believe the visual presentation of evidence is of particular value to a policymaker since the assessment of large (in this case, infinite) risk with small probabilities is something that may be poorly captured by simply reporting model averaged risk numbers.

When the policymaker has chosen the wrong theory, one again finds that for the hybrid case, the model-specific optimized rule strictly dominates the Taylor rule. This is interesting as it indicates that the failure to condition on lagged interest rates is a serious

deficiency of the Taylor for the hybrid case. For backwards models, one once again finds that there are 4 models for which the optimal rule induces instability, these are of course different from those found when the optimized rule is conditioned on the correct theory; the total posterior probability of these rules is .002.¹⁰ In addition, one finds that there are some models for which the Taylor rule outperforms the optimized rule even though the latter produces stability. There are 8 models of this type with posterior probability .08.¹¹ Interestingly, the models are generally those with longer lag lengths.

We next consider model averaging exercises that can reduce the information contained in Figure 1 down to a set of simple statistics. Table 4 reports risk ratios for model averaged risks. As noted above, the optimized rule produced instability for some backwards models, which would imply a value of infinity for the risk under the rule and would mean under an averaging calculation that the model averaged risk ratio and ratios of model averaged risk are both infinite. In order to produce nontrivial averaging calculations, for any model with infinite risk under the optimized rule we use risk values that produce a model-specific risk ratio of 5 and 20. Some authors do propose assigning a finite risk to unstable models (e.g. Del Negro and Schorfede (2004) who suggest the risk ratio 5 as a benchmark) while others assign infinite risk, (e.g. Levin and Williams (2003)). We also report (in column 5) the replacement values that will produce overall risk comparisons that render one indifferent between the Taylor and optimized rules; we remind the reader that this is only relevant for backwards models since the hybrid models are never unstable.

Our model averaging exercises uniformly provide support for the optimized rule over the Taylor rule. Interestingly, the optimized rule outperforms the Taylor rule even when the policymaker has erred in terms of choice of model class. This illustrates the value of allowing an interest rate rule to depend on lagged interest rates. We would also

⁹Letting (i, j, k) denote the model specification with i income lags in the IS equation, j income and k inflation lags in the Phillips curve equation, the models in which instability occurs are (1,1,1), (2,1,1), (3,1,1) and (4,1,1).

¹⁰The models which are unstable in this exercise are (1,2,4), (2,1,1), (3,4,4), and (4,1,1).

¹¹The models where the Taylor rule outperforms the model-specific optimized rule are (4,1,3), (4,1,4), (4,2,3), (4,2,4), (4,3,3), (4,3,4), (4,4,3) and (4,4,4).

note that one needs to assign model-specific risk ratios of about 65 for unstable models in order to conclude that the Taylor rule performed as well as the optimized rule.

Table 5 reports some summary statistics for our two exercises when the Taylor rule/optimized rule comparisons are restricted to models in which neither rule produces instability. Similar results hold for the cases for rules optimized on the highest posterior model in the correct class of models and rules optimized on the wrong class. One important feature of the Table is its demonstration that the relative risk between the two rules is extremely stable across the model specifications. As might have been expected given the findings in Table 4, this applies whether or not the rule is compared to the class whose maximum probability model was the basis of the rule. This implies, given our analysis of outcome dispersion for the Taylor rule, that the theory and model-specific optimized rule, modulo models where instability is induced, also has good properties in terms of producing stable (across models) outcome dispersion. In addition, it appears that assuming the backwards theory is true when it is not has low costs to a policymaker, at least in terms of comparisons to the Taylor rule.

These findings lead to the conclusion that virtues of the Taylor rule relative to an optimized rule derive from its ability to avoid producing model instability and otherwise that the optimized rules are uniformly better.

vi. outcome dispersion and action dispersion for optimal 3-variable rules

In our third exercise, we explore the sensitivity of optimal 3-variable rules to model choice. The idea in this work is to understand how the specification and associated risk of an optimal rule varies about specifications. Unlike the previous exercises, we do not specify a single rule and look at its behavior across models; each model is associated with its specific optimal rule. Table 6 presents information on the distribution of policy parameters and risk across models. The parameters were found with a grid search, with step size of 0.1, except for g_i for hybrid models in which a secondary grid search with steps of 0.02 was used because initially there was almost no variation across models to the first decimal place. Note that each column presents statistics across all 64 models.

To interpret the table, consider, for example, in the class of backward looking models, the minimum values presented in line (3). The minimum value of \tilde{g}_π of 2.9 need not have been found in the same specification that yielded the minimum value of \tilde{g}_y of 1.5, and neither of these specifications need have yielded the minimum value of risk R of 8.8.

We first consider the median values presented in line (5) of panel A. Consistent with previous literature such as Levin and Williams (2003), the hybrid model, which was solved treating expectations as model-consistent and thus forward looking, yields a lagged interest rate weight g_i that is higher than that for the backward model. In other respects the parameters are also congruent with earlier research. For example, in results not reported in the table we found that increasing λ_i shifts the distribution (across models) of the associated optimal g_i upwards; increasing λ_y also shifts the distribution of the associated optimal g_y upwards.

We have argued that there is relatively little outcome dispersion within a given class of models, at least if we focus on models with high posteriors. Table 7 illustrates that the same conclusion applies when we combine models from the two classes. We combine using a simple arithmetic average, as in Levin and Williams (2003). We do not weight by likelihood, as in much of the model averaging literature as well as our previous work (Brock, Durlauf, and West (2003) because, as noted above, the hybrid model explicitly was derived after a larger than usual amount of data mining. Panel A in Table 7 asks about outcome dispersion if we simply hold fixed the parameters at the values that are optimal for the likeliest backwards model (columns (1)-(3) in panel A) or likeliest hybrid model (columns (4)-(6)). Outcome dispersion is very small in columns (1)-(3); that is, a policy maker who is committed to using the parameters that are optimal for the likeliest backwards model is unlikely to be perturbed if he suddenly contemplates the possibility that hybrid model has a large element of truth as well. Outcome dispersion is, however, perceptible in columns (4)-(6).

The asymmetrical outcome results from the way we treated the two model classes. One could instead solve for parameters that are optimal given weights to each model. Results for this approach are given in panel B. The weight on the backwards model is

denoted θ ; results for $\theta=0$ and $\theta=1$ repeat results in Table 5.C and are given for reference. As one would expect, the policy parameters move smoothly as θ is varied. Unsurprisingly, action dispersion is small for \tilde{g}_π and moderate for \tilde{g}_y and g_i .

These tables may be complemented visually by graphs of the distributions of outcomes and actions across models. This is done in the set of pictures contained in Figure 2. As occurs in the reporting of objects such as impulse response functions from vector autoregressions, the visual reporting of outcome and action dispersion can suffer from a surfeit of information. We now turn to some suggestions on how these figures can be used by policymakers to inform decisions.

We first discuss action variance. Figure 2.A reports the different values of \tilde{g}_π , \tilde{g}_y and g_i that appear across the model-specific optimal rules in the backwards class. The panels depict visually the information on dispersion summarized in Table 6: there is a reasonable degree of dispersion across models with respect to \tilde{g}_π (in the sense of a support width of 1.0)¹², large dispersion with respect to \tilde{g}_y (support width of 1.8) and moderate dispersion for g_i (support width of .4). This implicitly means that the width of the support of the nonnormalized parameter g_π is about half that of the nonnormalized parameter g_y . Hence, policymakers can conclude that g_π is relatively insensitive to model specification. Within this variation, g_i is almost always greater than 0. This helps explain why the Taylor rule was generally inferior to three variable rules even when the latter was optimized on the wrong model. When one turns to the posterior weighted results, Figure 2.B, the main modification of these conclusions is that in some cases the supports of the parameters shrink when one focuses on those models whose posterior likelihoods are within 1/20 of the maximum posterior model. When one concentrates on these relatively likely models, one finds much smaller variation in \tilde{g}_π and \tilde{g}_y (measured by support width) than appears in Panel A. Interestingly, there is relatively less

¹²We focus on support width in our discussion of dispersion, information on standard deviations and interquartile ranges are available in Table 6 and yield qualitatively similar conclusions.

diminution of the support width of g_i for the relatively likely models. However, for the relatively likely models, g_i is always at least .1.

Similar results obtain for the hybrid model. Figure 2.C indicates that for this class, there is a larger support for the \tilde{g}_x and \tilde{g}_y parameters than in the backwards case (with support widths of 1.0 and 2.2 respectively). Compared with the backwards case, the variation in g_i is quite small, with a support width .1. When one turns to the posterior weighted results in Figure 2.D, one finds little reduction in support width when attention is restricted to the relatively likely models.

What conclusions might a policymaker draw? One conclusion is that conditioning on lagged interest rates is a robust feature of optimal policies. A second conclusion is that if one conditions policy on the hybrid class, the interest rate parameter in a three-variable interest rule of the form (17) is insensitive to lag length specification whereas in other contexts, the optimal rule parameters can vary substantially across specifications.

We next consider the dispersion of risk for the backwards models and the hybrids and compare. An examination of dispersion in risk across all the models for backwards and we see clustering at around 10, 15, and 25 whereas for hybrids risk is essentially clumped around 6 or 7 (lower right panels of Figures 2.A and 2.C). A policymaker who believed strongly in a backward looking world will want to proceed cautiously and look closely at what is generating this dispersion in risk. Perhaps most of the models generating the wide dispersion have low posterior probability. If one then examines the posterior weighted dispersion plot in the lower right hand panel of Figure 2,B it is evident that the risk clumping around 15 and 25 is generated by models with very low posterior probability. The policymaker may now be quite relieved and simply concentrate on managing the cluster of models whose risk clumps around 10. Further information is provided by focusing on relatively likely models. This restriction would lead a policymaker to concentrate attention on managing in a world dominated by the four models that clearly stand out on the plot as having the bulk of the posterior probability.

For the hybrid class, risk dispersion is very narrow in comparison to the backwards looking models. Whatever dispersion is observed is reduced further when computed with posterior weights and clumps around about 7.4 when one focuses on the

relatively likely models. This indicates substantial robustness for the optimal rules for hybrids.

This type of discussion, in which one compares the plots of unweighted and posterior weighted results, with further attention to the relatively likely models, enables a policymaker to get a good overview of the risk dispersion it must face and whether it is caused by models that are supported by the data in the sense that their posterior weights are relatively high. As such, this discussion suggests potential ways of dealing with critiques of the minimax criterion as being too fragile in the sense that it is influenced far too much by models that have extremely small probabilities either in a posterior sense or in some judgmental sense. The performance of the minimax criterion might be improved by applying it to a data determined “trimmed” subset of the possible models, e.g. the subset consisting of the 1/20 of the likeliest that we have employed. This same argument might be applied with profit to any criterion that can be unduly influenced by models with small “believability” whether believability is measured by posterior probability or some other method.

vii. patterns¹³

We finally note that there exists an interesting pattern that relates model complexity (in our context, length of lags) and the policy parameters. As indicated in Figure 3, while there is weak association between the total complexity of a model and the associated parameters, relatively strong patterns emerge when one considers IS curve complexity (the number of lags in eq. (18)) and with Phillips curve complexity (the number of lags in (19)). For backward looking models, the magnitude of g_i decreases in IS complexity but increases in PC complexity. The magnitude of g_y is decreasing with respect to both IS and PC complexity.

Different patterns emerge for the hybrid models. For this model class, one finds that g_i increases in IS complexity. The g_y parameter is increasing in both IS and PC

¹³Giacomo Rondina has greatly helped us in identifying these patterns.

complexity. These patterns are the opposite of what holds for the backwards-looking model.

These systematic pattern relationships for backwards-looking and hybrid models suggest some interesting avenues for future research. One question is whether these patterns are sensitive to the choices of λ_t and λ_y . A second broader question concerns the existence of patterns for more complex versions of the policy rule, such as rules which allow for policy lags beyond a single period. Brock and Durlauf (2004b) shows how, when control is costless, as the number of lags in the policy rule is allowed to become arbitrarily long, the variation in the state variables of a system is reduced to the variation of the i.i.d. drivers of the system. We conjecture that this also holds when the cost of control is small, i.e. λ_t is much smaller than λ_y in the current context. Hence a system in which the number of control parameters is highly restricted will not be able to achieve the Brock and Durlauf (2004b) reduction to fundamental i.i.d. shocks. The more complex the state equation, the greater the implicit restrictions on a simple rule such as (17) and hence the greater the “strain” on the rule to achieve this limit. We conjecture that there is something analogous to a Le Chatelier principle that produces a relationship between the Taylor parameters as the complexity of the state equation increases.

IV. Interpretation

In this section we consider some interpretations of our results in the context of an abstract dynamic system. We consider the backwards-looking class of models. This system is one dimensional, unlike the system we have studied empirically; we employ a one dimensional system as closed forms solutions are straightforward to develop for this case whereas for higher dimensional cases they are far more complicated and lead to a loss of intuition. Let x_t denote the state of the system and u_t denote the scalar control available to the policymaker. The state evolves according to

$$x_t = a(L)x_{t-1} + b(L)u_{t-1} + \xi_t \quad (28)$$

where the Wold representation of ξ_t is denoted

$$\xi_t = w(L)v_t \quad (29)$$

We assume that $w(L)$ is invertible. A policymaker has access to linear feedback rules of the form

$$u_{t-1} = -g(L)x_{t-1} \quad (30)$$

and chooses a feedback rule in order to minimize

$$Ex^2 + \lambda Eu^2 \quad (31)$$

We now consider a special case of this model: $\lambda = 0$, and $w(L) = 1$. For this class of models, the optimal choice¹⁴ of $g(L) = g^*(L)$ will fulfill

$$g^*(L) = \frac{a(L)}{b(L)} = 1 + d(L) \quad (32)$$

Eq. (32) is useful because it illustrates the basic Taylor principle for stabilization policy. To see this, consider the special case $a(L) = a \in (0,1)$ and $b(L) = b$, so that $g^*(L) = 1 + d = 1 + \frac{a-b}{b}$. Relative to the model in Section III, one can equate x_t with inflation and u_t with the nominal interest rate. The Taylor principle is $g_\pi > g_y$, so that inflation innovations get greater weight than output innovations. By analogy, we have

¹⁴This finding is standard; we refer the reader to Brock and Durlauf (2004b) for a rigorous development of necessity and sufficiency arguments for models of this type.

the same tendency to react relatively strongly to inflation. For our special case, the magnitude of the feedback from last period's inflation to today's nominal interest rate is $\frac{a-b}{b}$. If $\frac{a-b}{b} > 0$ the feedback is more than one to one. This seems an empirically plausible case given the high persistence in the inflation series. One could also argue that Friedman's classic (1948) concern about long and variable lags is interpretable as suggesting that the feedback polynomial $b(L)$ is not that persistent.

This model may be used to illustrate the concepts of outcome dispersion and action dispersion we have described in Section III.vi. In doing this, we will ignore parameter uncertainty. We first consider the case where the optimal policy is not constrained in terms of numbers of lags. Let the model space M be defined as

$$M = \{a(L, m), b(L, m)\} \quad (33)$$

Where $a(L, m)$ and $b(L, m)$ denote model-specific lag polynomials. In our analysis, we considered a set of 64 different models for the model space (33). Each model is associated with a distinct fundamental driver $v_{m,t}$ with variance $\sigma_{v_m}^2$.

If $w(L)=1$, outcome dispersion is generated by cross-section variation in $\sigma_{v_m}^2$, recalling our assumption that the lag length for the policy rule is not constrained. The model-specific optimal rule eliminates all dependence in the state. Action dispersion in this case refers to the variance of g_j^* , the coefficient associated with L^j in $g^*(L)$. For model m , which is a joint specification of $a(L, m)$ and $b(L, m)$, there will be an associated $g^*(L)$, hence g_j^* may vary across models even if the outcome dispersion does not. The variance of g_j^* can be written as

$$\text{var}(g_j^*) = \text{var}(1 + d(L, m)_j) \quad (34)$$

In this expression $(r(L))_j = r_j$.

These calculations assume that a policymaker may choose any lag length for the feedback rule. One may ask similar questions about outcome and action dispersion when policymakers are required to choose rules with restrictions on lag length; in fact many of the “simple” rules that have been considered in recent monetary research, of which the Taylor rule is a leading example, in fact do this. From the perspective of model uncertainty in lag structure, these simple rules run the risk of being unable to counter longer-run feedbacks.

To understand the costs of overly simple rules, we consider the case $w(L) = 1$ and $b = 1$. Suppose that the true model is one where the lag structure for $a(L)$ contains N lags. If one were to consider a sequence of optimal rules, in which the k 'th rule is constrained to only have k lags, then it is easy to see that the value of Ex^2 obtainable with a k -lag rule is decreasing (in k) and will, when $k = N$ equal σ_v^2 . This simple logic is suggestive of the factors that will determine the outcome dispersion for a model space of the form $M = \{a(L, m), b\}$. If the set of possible policy rules allows for lags lengths up to N , then the minimum outcome dispersion may be obtained for every model in M .

This basic argument has an important implication for outcome dispersion and model uncertainty: outcome dispersion will decline to 0 as the number of lags in the policy rule space increases. Conversely, if one defines a *complexity gap* as the difference between the number of lags in the state equation and the number of lags in the policy rule, one would expect the estimated risk to be increasing in this gap. The dispersion plots for minimum risk in Figure 2 appear to possess this property. This is so because we optimized over parameters for the single-lag structure where the total number of lags in the behavioral equations increases from 3 to 12 as we move across the model space.

V. Conclusions

In this paper, we have attempted to outline some basic principles for incorporating model uncertainty into the reporting of policy evaluation exercises. We have argued that the policy analysis should not be done conditional on a specific model but rather should reflect model uncertainty. This leads to model averaging methods that treat model specification as an unobservable in a way parallel to any other type of unknown in data analysis. We have applied these ideas to some monetary policy exercises. These exercises suggest that the Taylor rule has good robustness properties. These analyses also suggest some ways to visualize the role of model uncertainty which may facilitate communication with policymakers.

To be clear, our analysis really only scratches the surface of the many questions that arise when model uncertainty is incorporated into policy exercises. One important question is how to operationalize our approach to richer model spaces, such as spaces which incorporate various types of learning and nonlinearity. Another question concerns the appropriate specification of prior probabilities on model spaces for macroeconomic contexts such as monetary policy evaluation. Perhaps most important, our analysis describes uncertainty for a fixed model space. Since progress in economic research should have the effect of expanding the space over time, this expansion should be incorporated into any decision problem. It might well also be the case that the choice of rules should reflect the implications of a rule for how information about a model space is produced. All of these questions suggest that model uncertainty research should prove an active area of study.

Table 1
Parameter Estimates for Models with Highest Posterior

A. IS curve

	α_{y1}	α_{y2}	α_{y3}	α_r	\bar{R}^2	D.W.	s.e.
Backward	1.12 (0.09)	-0.04 (0.13)	-0.20 (0.08)	0.07 (0.03)	0.89	2.03	0.78
Hybrid	1.10 (0.09)	-0.21 (0.08)	n.a.	0.13 (0.03)	0.89	2.06	0.77

B. Phillips curve

	β_{y1}	$\beta_{\pi1}$	$\beta_{\pi2}$	$\beta_{\pi3}$	β_0	\bar{R}^2	D.W.	s.e.
Backward	0.16 (0.04)	0.69 (0.08)	0.01 (0.10)	0.30 (0.08)	n.a.	0.83	2.09	1.07
Hybrid	0.14 (0.04)	1.00	n.a.	n.a.	0.32 (0.07)	0.83	2.11	1.07

Notes:

1. Panel A presents estimates of equation (19), panel B estimates of (20). Constant terms were included in all regressions. The backward and hybrid models differ in their treatment of expected inflation, as explained in the text.

2. In panel A, the output gap is the dependent variable, α_{yj} is the coefficient on output gap at lag j , α_r the coefficient on the annual real interest rate. In panel B, inflation is the dependent variable, β_{y1} is the coefficient on y_{t-1} , $\beta_{\pi j}$ the coefficient on inflation at lag j , β_0 the coefficient on a survey measure of expected annual inflation.

3. The data are quarterly. The sample of 131 observations is 1970:2-2002:4. Inflation is the annualized change in the GDP deflator; the output gap is computed from real GDP and the CBO estimate of potential GDP; the interest rate is the average Federal funds rate.

Table 2**Relative Likelihood P**

	<u>Backward</u>	<u>Hybrid</u>
(1)Minimum P	1×10^{-7}	2×10^{-6}
(2)Q1 P	2×10^{-5}	2×10^{-4}
(3)Median P	3×10^{-4}	1×10^{-3}
(4)Q3 P	2×10^{-3}	9×10^{-3}
(5)Maximum P	0.30	0.30
(6)No. models with $P > (\max P)/20$	8	13
(7)Sum of P for models with $P > (\max P)/20$	0.92	0.89
(8)Sum of P for models in top quartile	0.98	0.93
(9)Sum of P	1.0	1.0

Notes:

1. Let \hat{L}_m be the BIC-adjusted likelihood for model m . Then

$$P_m = \frac{\hat{L}_m}{\sum_{m \in C} \hat{L}_m}$$

where the summation runs over the 64 models in a given class (backwards or hybrid). As indicated in line (9), by construction $\sum_m P_m = 1$.

Table 3
Risk and Model Uncertainty for Original Taylor Rule

Taylor Rule: $g_x=1.5, g_y=0.5, g_i=0$

	A. All Models	
	<i>Backwards</i>	<i>Hybrids</i>
(1)Mean	30.0	25.6
(2)Std. Dev.	10.9	7.1
(3)Minimum	17.5	12.5
(4)Q1	20.2	20.1
(5)Median	26.3	26.0
(6)Q3	36.5	30.8
(7)Maximum	51.6	44.7
(8)Post. Weighted Av.	22.0	23.6

B. Models with High Posterior Probability

(1)Minimum	19.1	15.2
(2)Maximum	23.2	31.9

Notes:

1. This table presents information on the distribution across the 64 models in a given class (backward looking or hybrid) of risk R when monetary policy follows Taylor (1993) rule given in the header of the table.
2. The risk function is given in (18), $R = var(\pi_\infty) + \lambda_y var(y_\infty) + \lambda_i var(\Delta i_\infty)$, for $\lambda_y=1.0$ and $\lambda_i=0.1$.
3. In panel B, “high” posterior probability is defined as having a BIC adjusted likelihood at least 1/20 of the model with the highest BIC adjusted likelihood.

Table 4

Ratio of Risk from Taylor (1993) Rule to Risk from Optimized Rules

Class of models (Optimal 3 Variable Rule used)	(1) Omitting Unstable Models	(2) Ratios for Unstable = 5.0	(3) Ratios for Unstable = 20.0	(4) Equivalent Ratio for Unstable
(1) Backward (Optimal Backward)	0.55	0.56	0.61	135
(2) Backward (Optimal Hybrid)	0.75	0.77	0.80	107
(3) Hybrid (Optimal Hybrid)	0.32	0.32	0.32	n.a.
(4) Hybrid (Optimal Backward)	0.38	0.38	0.38	n.a.

Notes:

1. This table presents the posterior weighted average ratios of risk R when monetary policy follows the Taylor (1993) rule to risk when monetary policy follows certain optimized rules. These optimized rules set the interest rate i as in (17), $i_t = g_\pi \pi_t + g_y y_t + g_i i_{t-1}$. The parameters g_π , g_y and g_i are chosen to minimize risk R given the estimates of the IS and Phillips curves presented in Table 1 above. Denote risk from Taylor (1993) rule as \hat{R}_T and risk from an optimized rule as \hat{R}_O , then the posterior weighted average ratio is:

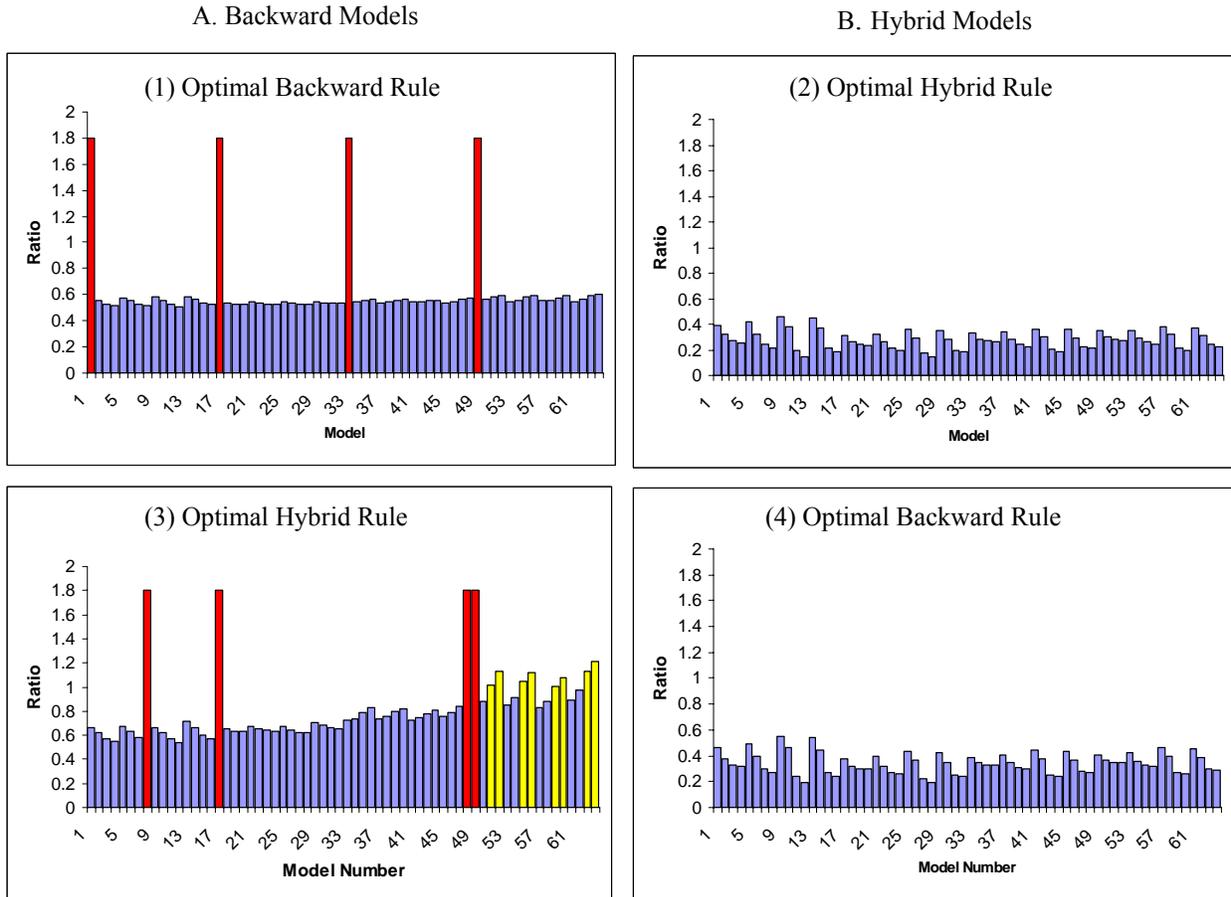
$$\Sigma_{m \in C} P_m (\hat{R}_O / \hat{R}_T)$$

2. Lines (1) and (2) report the average ratio for the backward models using the optimized rule for the likeliest backward model (in line (1)) and the optimized rule for the likeliest hybrid model (in line (2)). Similarly, lines (3) and (4) report the average ratio for the hybrid models using the optimized rule for the likeliest hybrid model (in line (3)) and the optimized rule for the likeliest backward model (in line (4)).

3. Column (1) reports average ratios when unstable models are omitted from the calculation. Column (2) and (3) report average ratio when the risk assigned to unstable models is so that their ratio to Taylor (1993) is 5.0 (column (2)) and 20.0 (column (3)).

4. Column (4) reports the ratios that have to be assigned to unstable models in order to obtain an average ratio equal to 1.0, meaning that the Taylor (1993) rule is equivalent to the optimized rule when considering posterior weighted averages.

Figure 1
Ratios of Risk for Optimal Policy Rules over Original Taylor Rule



Notes:

1. This figure presents the ratio of risk R when monetary policy follows the Taylor (1993) rule to risk when monetary policy follows certain optimized rules. These optimized rules set the interest rate i as in (17), $i_t = g_\pi \pi_t + g_y y_t + g_i i_{t-1}$. The parameters g_π , g_y and g_i are chosen to minimize risk R given the estimates of the IS and Phillips curves presented in Table 1 above.

2. The policy rules are:

Original Taylor Rule: $g_\pi=1.5, g_y=0.5, g_i=0$

Optimized 3 Variable Backward: $g_\pi=3.2, g_y=2.1, g_i=0.2$

Optimized 3 Variable Hybrid: $g_\pi=3.2, g_y=4.7, g_i=0.55$.

3. In panels A the denominator is the risk R obtained applying Taylor (1993) rule to backward models: in (1) the numerator is the risk obtained using the optimized rule for the likeliest backward model; in (3) the numerator is the risk obtained using the optimized rule for the likeliest hybrid model.

In the panels B the denominator is the risk R obtained applying Taylor (1993) rule to hybrid models: in panel (2) the numerator is the risk obtained using the optimized rule for the likeliest backward model; in (4) the numerator is the risk obtained using the optimized rule for the likeliest hybrid model. In either case, the ratios are computed using the IS and Phillips curve estimates of 64 models in each class. See Appendix 1 for a mapping of the model numbers to details of specification of IS and Phillips curves.

4. A ratio less than one means that the optimized rule delivers less risk than did the original (1993) Taylor rule. If the optimized rule led to instability, the ratio is truncated at 1.8. The original Taylor rule did not lead to instability in any models.

Table 5**Risk Distributions Across Models:**

	A. All Models					
	Backward			Hybrid		
	Risk Original Taylor	Risk Optimized 3 Variable	Ratio (Optimized over Original)	Risk Original Taylor	Risk Optimized 3 Variable	Ratio (Optimized over Original)
(1)Mean	30.0	16.0	0.55	25.6	6.7	0.28
(2)Std. Dev.	10.9	5.7	0.022	7.1	0.5	0.073
(3)Minimum	17.5	9.2	0.51	12.5	5.7	0.15
(4)Q1	20.2	11.2	0.54	20.1	6.2	0.22
(5)Median	26.3	13.1	0.55	26.0	6.7	0.27
(6)Q3	36.5	18.7	0.56	30.9	6.9	0.33
(7)Maximum	51.6	28.1	0.61	44.7	7.6	0.46

B. Models with High Posterior Probability

	Backward (8 models)			Hybrid (13 models)		
(1)Minimum	19.1	10.3	0.53	15.2	6.3	0.23
(2)Maximum	23.2	13.1	0.59	31.9	7.6	0.42

C. Model with Highest Posterior Probability

	Backward			Hybrid		
(1)	23.2	12.9	0.55	24.1	7.6	0.32

Notes:

1. The policy rules are:

Original Taylor Rule: $g_{\pi}=1.5$, $g_y=0.5$, $g_i=0$ Optimized 3 Variable Backward: $g_{\pi}=3.2$, $g_y=2.1$, $g_i=0.2$ Optimized 3 Variable Hybrid: $g_{\pi}=3.2$, $g_y=4.7$, $g_i=0.55$.2. The assumed monetary policy rule is given in (17), $i_t = g_{\pi}\pi_t + g_y y_t + g_i i_{t-1}$. The risk function is given in (18), $R = \text{var}(\pi_{\infty}) + \lambda_y \text{var}(y_{\infty}) + \lambda_i \text{var}(\Delta i_{\infty})$, for $\lambda_y=1.0$ and $\lambda_i=0.1$.

3. In panel B, “high” posterior probability is defined as having a BIC adjusted likelihood at least 1/20 of the model with the highest BIC adjusted likelihood.

4. The regression estimates for models with the highest probabilities are given in Table 1.

Table 6

Distribution of Optimal Policy Parameters and Risks Across Models

A. All Models

	Backward				Hybrid			
	$g_\pi/(1-g_i)$	$g_y/(1-g_i)$	g_i	R	$g_\pi/(1-g_i)$	$g_y/(1-g_i)$	g_i	R
(1)Mean	3.4	2.4	0.2	16.3	3.2	5.0	0.56	6.6
(2)Std. Dev.	0.3	0.5	0.1	5.9	0.4	0.6	0.03	0.6
(3)Minimum	2.9	1.5	0.0	8.8	2.3	3.7	0.51	5.5
(4)Q1	3.2	2.0	0.1	11.3	2.7	4.6	0.53	6.2
(5)Median	3.4	2.3	0.2	14.6	3.0	5.0	0.57	6.8
(6)Q3	3.6	2.7	0.3	20.2	3.1	5.5	0.57	7.0
(7)Maximum	3.9	3.3	0.4	27.3	3.3	5.9	0.59	7.7

B. Models with High Posterior Probability

	Backward (8 models)				Hybrid (13 models)			
(1)Minimum	3.0	2.0	0.1	10.2	2.6	3.7	0.51	6.2
(2)Maximum	3.5	2.6	0.3	13.1	3.8	5.7	0.59	7.8

C. Model with Highest Posterior Probability

	Backward				Hybrid			
(1)	3.2	2.1	0.2	12.9	3.2	4.7	0.55	7.6

Notes:

1. This table presents information on the distribution across the 64 models in a given class (backward looking or hybrid) of monetary policy parameters g_π , g_y and g_i that yielded minimum risk R . The values were found by grid search over g_π , g_y and g_i .
2. The assumed monetary policy rule is given in (17), $i_t = g_\pi \pi_t + g_y y_t + g_i i_{t-1}$. The risk function is given in (18), $R = \text{var}(\pi_\infty) + \lambda_y \text{var}(y_\infty) + \lambda_i \text{var}(\Delta i_\infty)$, for $\lambda_y=1.0$ and $\lambda_i=0.1$.
3. In panel B, “high” posterior probability is defined as having a BIC adjusted likelihood at least 1/20 of the model with the highest BIC adjusted likelihood.
4. The regression estimates for models with the highest probabilities are given in Table 1.

Table 7

Optimal Policy When Combining Hybrid and Backwards Models

A. Policy parameters are held fixed at levels optimal for likeliest model in a given class

(1)	(2)	(3)	(4)	(5)	(6)
Held fixed at backwards level			Held fixed at hybrid level		
R_b^*	R_h	$.5R_b^* + .5R_h$	R_b	R_h^*	$.5R_b + .5R_h^*$
12.9	9.0	11.0	18.3	7.6	12.9

B. Optimization over a weighted average of a single backwards and single hybrid model

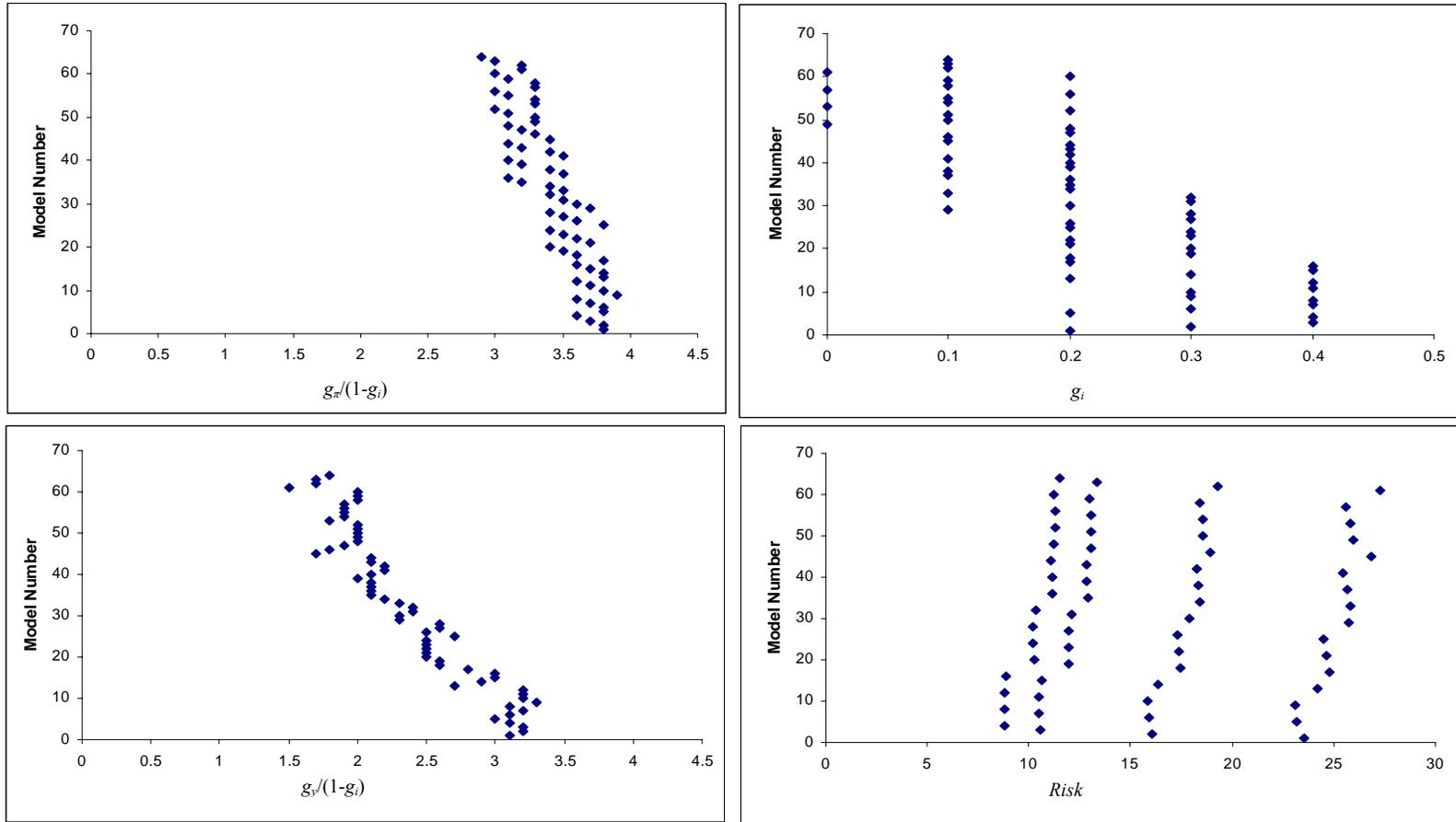
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Backwards	$g_\pi/(1-g_i)$	$g_y/(1-g_i)$	g_i	R_b	R_h	$\theta R_b + (1-\theta)R_h$
Weight (θ)						
0	3.2	4.7	0.55	18.3	7.6	7.6
0.25	3.1	3.2	0.41	13.9	7.9	9.4
0.5	3.2	2.7	0.31	13.2	8.3	10.7
0.75	3.2	2.3	0.25	12.9	8.7	11.8
1.0	3.2	2.1	0.2	12.9	9.0	12.9

Notes:

1. Let $R_b^* = 12.9$ and $R_h^* = 7.6$ denote risk that obtains when the model that is likeliest within a given class of models is used, see Table 6.C. In column (2) of panel A, R_h denotes the risk that obtains for the likeliest hybrid model (parameter estimates in Table 1) when the policy parameters are held fixed at the values that lead to R_b^* . By construction, R_h is at least as large as R_h^* . In column (4) of panel A, R_b is similarly computed, using backwards model estimates presented in Table 1 and hybrid policy parameters presented in Table 6.C.

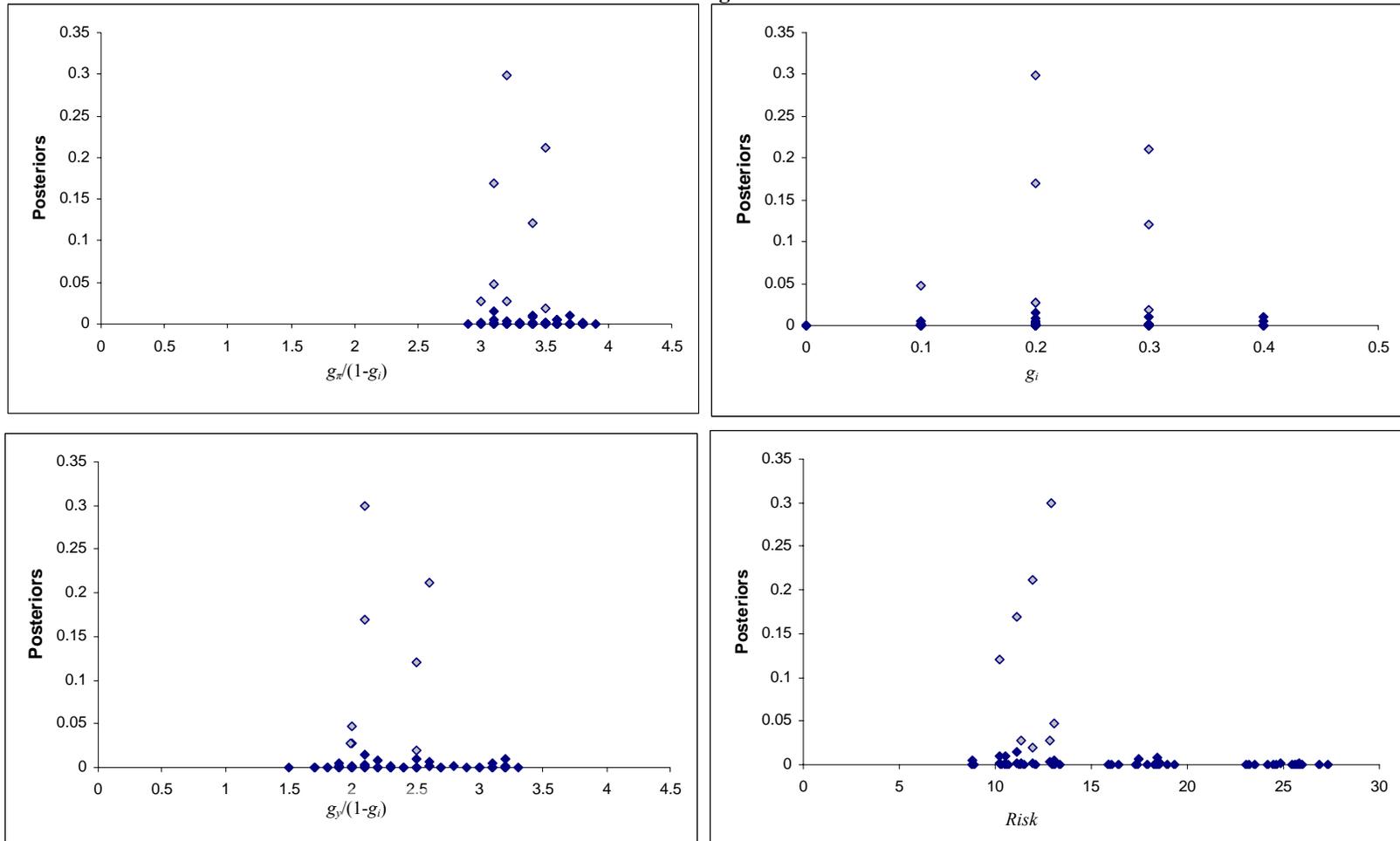
2. Panel B present parameters that are optimal when the risk function is the indicated arithmetic average of backwards and hybrid models. Risk for $\theta=0$ and $\theta=1.0$ corresponds to what is called R_h^* and R_b^* in panel A.

Figure 2 Outcome and Action Dispersion
Parameters Results for Backward Models
A.Non-Weighted Results



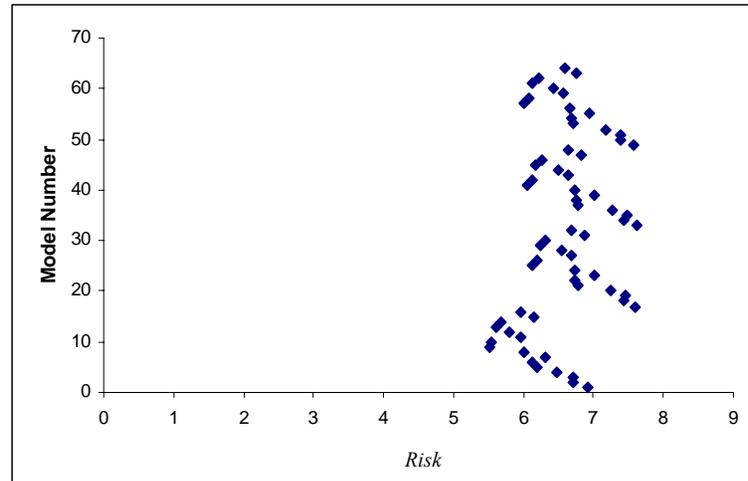
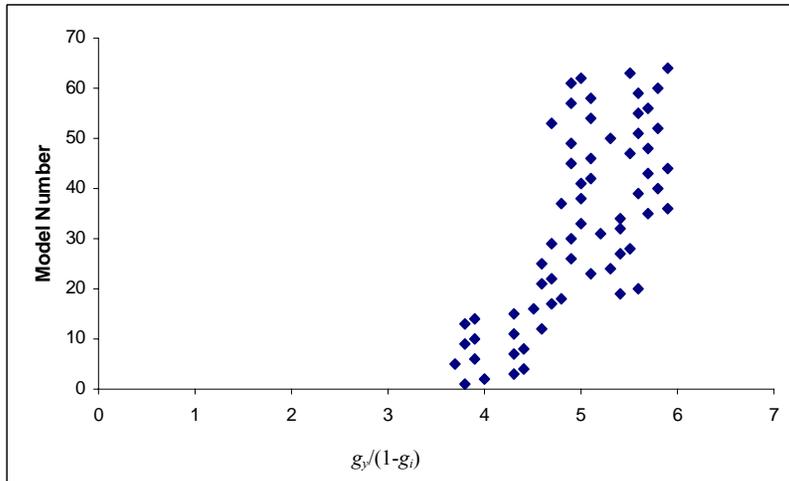
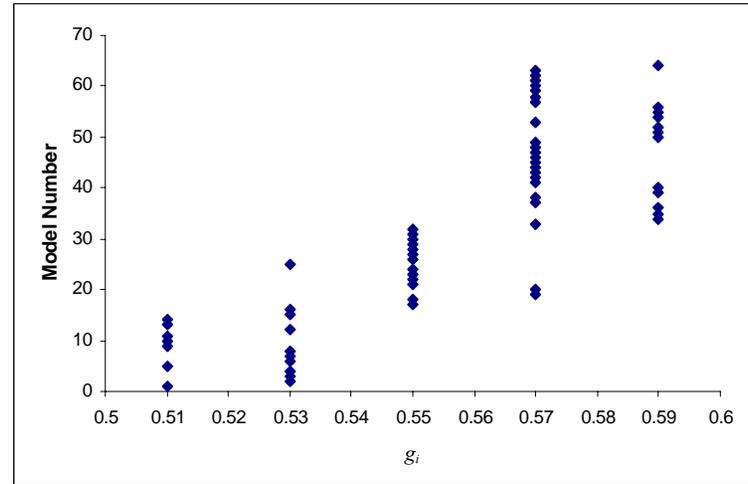
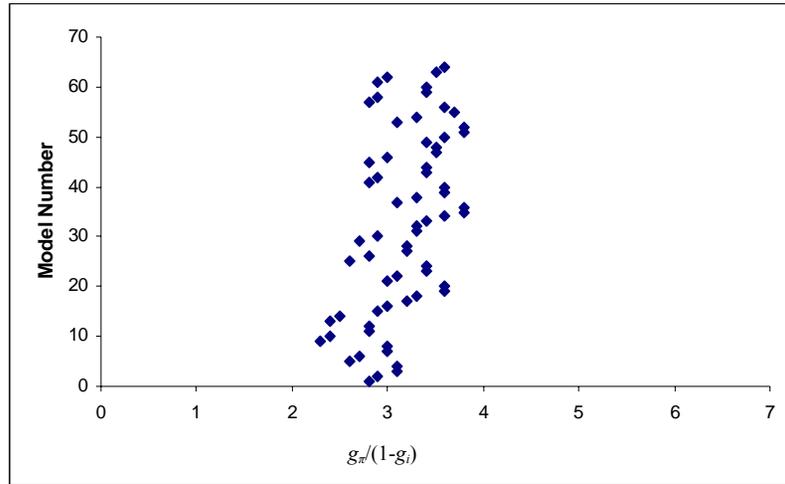
Notes: The two top panels and the left bottom panel report the value of the optimal policy parameter for each model (indexed by model's number, see Appendix 1). The right bottom panel reports the values of the minimum risk for each model corresponding to the optimal parameters found. Risk R is calculated using preference values: $\lambda_y = 1.0$ and $\lambda_i = 0.1$, where $R = \text{var}(\pi_{\infty}) + \lambda_y \text{var}(y_{\infty}) + \lambda_i \text{var}(\Delta i_{\infty})$.

B. Posterior Weighted Results



Notes: Panels B report the same results as Panels A concerning the parameter values and the minimum risk. This time they are plotted against the relative BIC adjusted relative likelihood of each model. The light shaded dots refer to models having a BIC adjusted likelihood at least 1/20 of the model with the highest BIC adjusted likelihood.

Parameters Results for Hybrid Models
C. Non-Weighted Results



D. Posterior Weighted Results

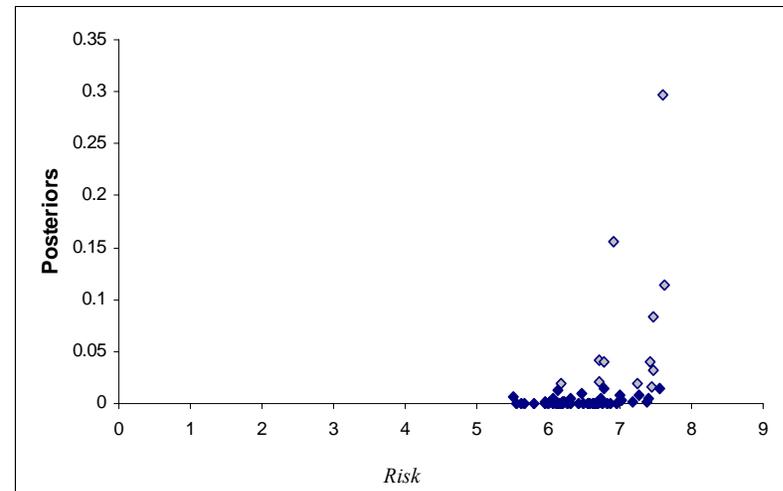
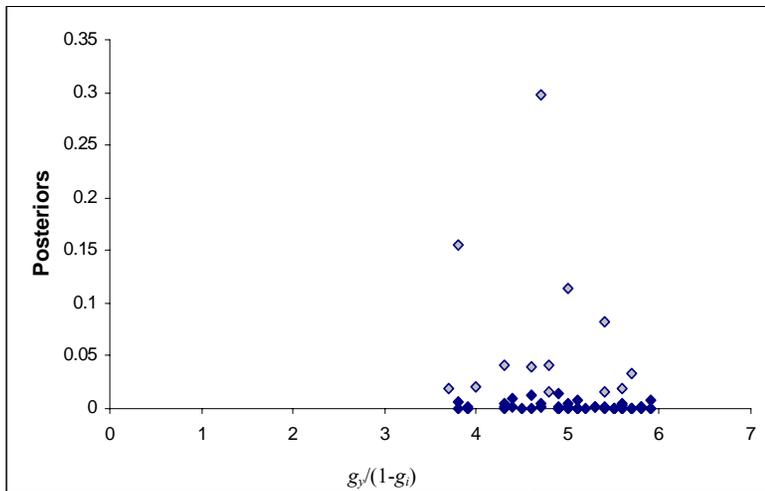
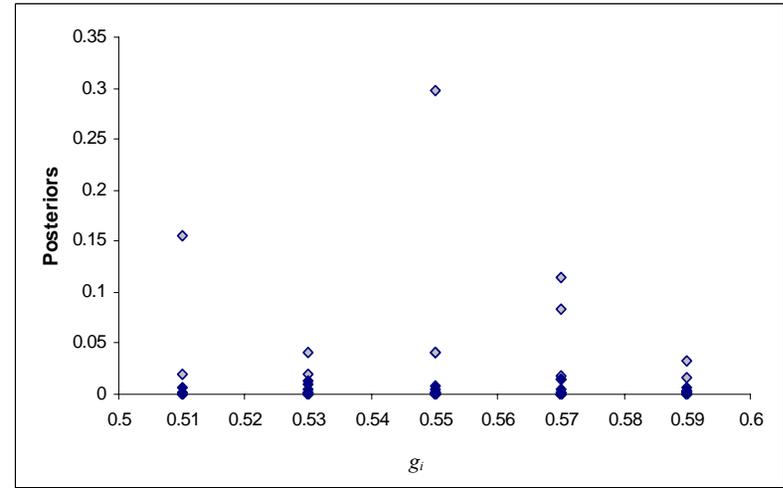
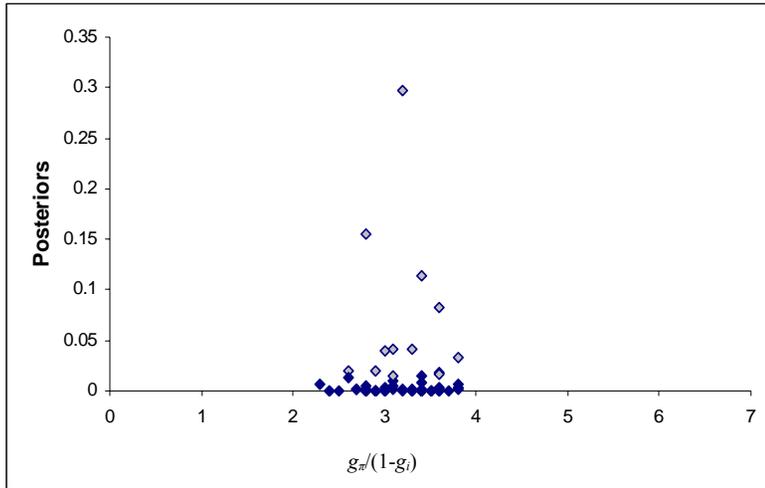
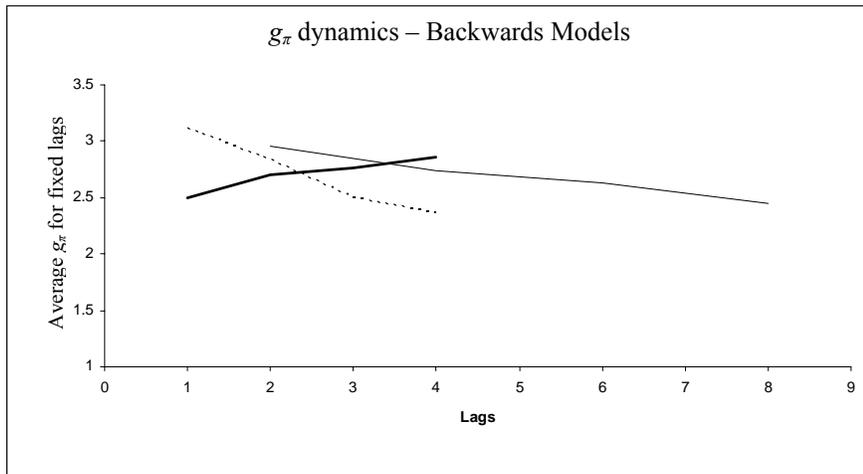
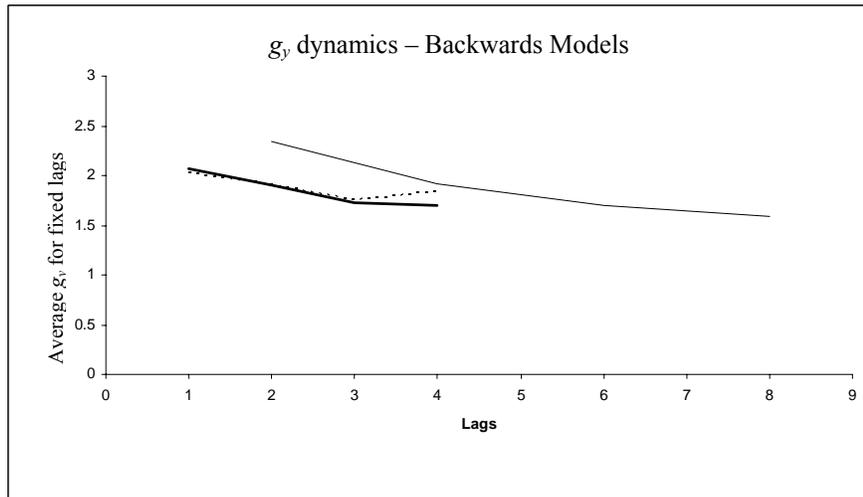
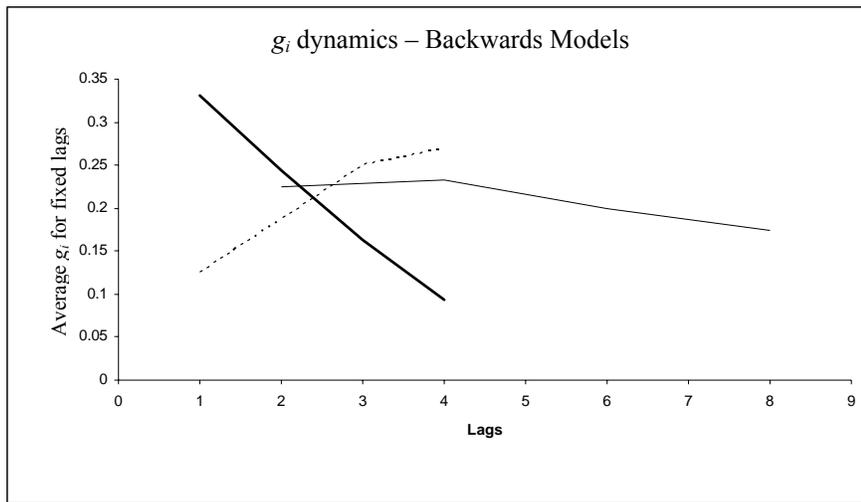
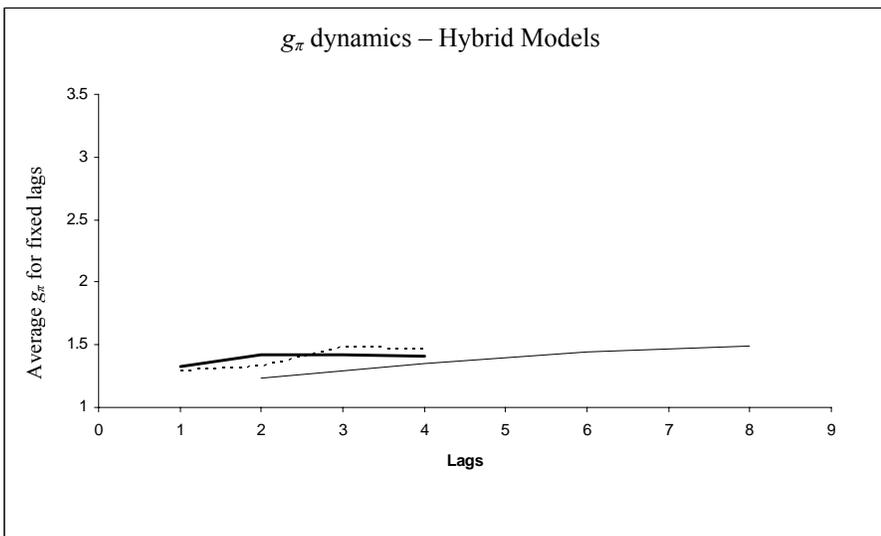
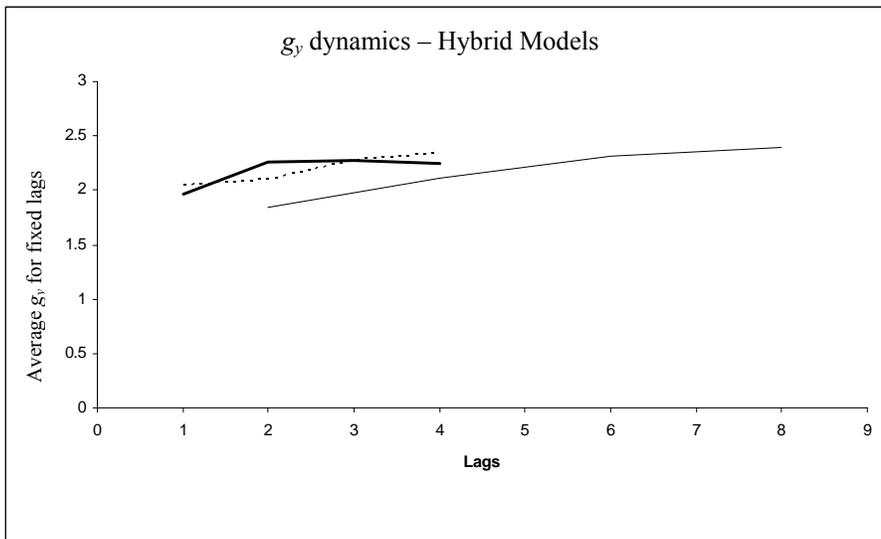
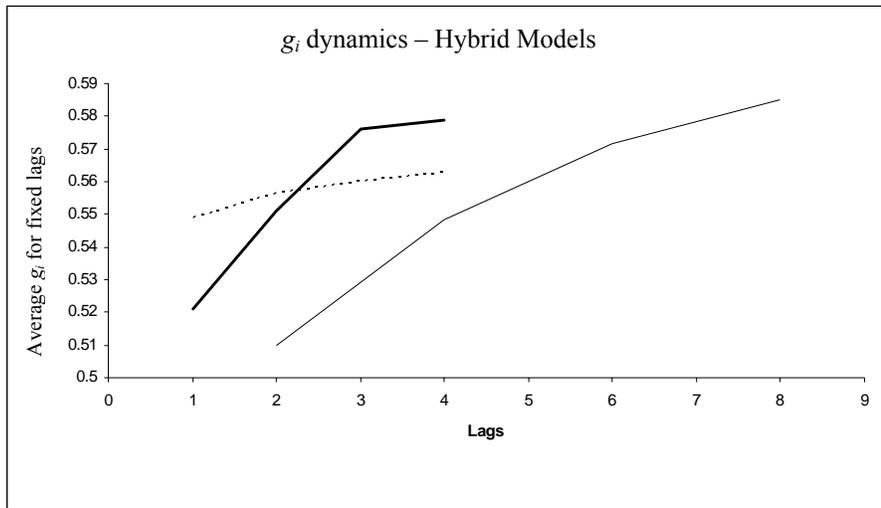


Figure 3
Model Complexity and Parameter Relationship



— y lags in IS - - - π lags in PC — Total lags (y in IS and π in PC)



y lags in IS
 π lags in PC
 Total lags (y in IS and π in PC)

Appendix I

This appendix maps the model numbers used in Figure 1 and 2 into details of specifications of the IS and Phillips curves. For each model number running from 1 to 64, three numbers are presented. These are: number of lags of y in IS curve; number of lags of y in Phillips curve; number of lags of π in Phillips curve. For example, model 25 had 2 lags of y in the IS curve, along with 3 lags of y and 1 lag of π in the Phillips curve.

Index for Model Space

Model Number	Specification y lags in IS, y lags and π lags in PC						
1	1 1 1	17	2 1 1	33	3 1 1	49	4 1 1
2	1 1 2	18	2 1 2	34	3 1 2	50	4 1 2
3	1 1 3	19	2 1 3	35	3 1 3	51	4 1 3
4	1 1 4	20	2 1 4	36	3 1 4	52	4 1 4
5	1 2 1	21	2 2 1	37	3 2 1	53	4 2 1
6	1 2 2	22	2 2 2	38	3 2 2	54	4 2 2
7	1 2 3	23	2 2 3	39	3 2 3	55	4 2 3
8	1 2 4	24	2 2 4	40	3 2 4	56	4 2 4
9	1 3 1	25	2 3 1	41	3 3 1	57	4 3 1
10	1 3 2	26	2 3 2	42	3 3 2	58	4 3 2
11	1 3 3	27	2 3 3	43	3 3 3	59	4 3 3
12	1 3 4	28	2 3 4	44	3 3 4	60	4 3 4
13	1 4 1	29	2 4 1	45	3 4 1	61	4 4 1
14	1 4 2	30	2 4 2	46	3 4 2	62	4 4 2
15	1 4 3	31	2 4 3	47	3 4 3	63	4 4 3
16	1 4 4	32	2 4 4	48	3 4 4	64	4 4 4

Bibliography

Avramov, D. (2002). "Stock Return Predictability and Model Uncertainty," *Journal of Finance*, 64, 423-458.

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

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

Brock, W. and S. Durlauf. 2001. "Growth Empirics and Reality." *World Bank Economic Review*, 15, 229-272.

Brock, W. and S. Durlauf. 2004a. "Elements of a Theory of Design Limits to Optimal Policy," *The Manchester School*, 72, Supplement 2, 1-18.

Brock, W. and S. Durlauf. 2004b. "Local Robustness Analysis: Theory and Application." Mimeo, University of Wisconsin.

Brock, W. and C. Hommes. 1997. "A Rational Route to Randomness." *Econometrica*, 65, 5, 1059-1096.

Brock, W., S. Durlauf and K. West. 2003. "Policy Analysis in Uncertain Economic Environments (with discussion)." *Brookings Papers on Economic Activity*, 1, 235-322, 2003.

Brown, P., M. Vannucci, and T. Fearn. 1998. "Multivariate Bayesian Variable Selection and Prediction." *Journal of the Royal Statistical Society, series B*, 60, 627-41.

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

Cogley, T. and T. Sargent. 2004. "The Conquest of U.S. Inflation: Learning and Robustness to Model Uncertainty." Mimeo, University of California at Davis.

Del Negro, M. and F. Schorfede. 2004. "Policy Predictions if the Model Doesn't Fit." Mimeo, University of Pennsylvania.

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

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

- Epstein, L. and T. Wang. 1994. "Intertemporal Asset Pricing Behavior Under Knightian Uncertainty." *Econometrica*, 62, 283-322.
- Fernandez, C., E. Ley and M. Steel. 2001. "Model Uncertainty in Cross-Country Growth Regressions." *Journal of Applied Econometrics*, 16, 5, 563-76.
- Friedman, M. 1948. "A Monetary and Fiscal Framework for Economic Stability." *American Economic Review*, 38, 245-264.
- Galí, J. and M. Gertler. 1999. "Inflation Dynamics: A Structural Econometric Model." *Journal of Monetary Economics*, 44, 195-222.
- Garratt, A., K. Lee, M. H. Pesaran, and Y. Shin. 2003. "Forecasting Uncertainties in Macroeconometric Modelling: An Application to the UK Economy." *Journal of the American Statistical Association*, 98, 464, 829-838.
- Giannoni, M. 2001. "Robust Optimal Monetary Policy in a Forward-Looking Model with Parameter and Shock Uncertainty," mimeo, Federal Reserve Bank of New York.
- Giannoni, M. 2002. "Does Model Uncertainty Justify Caution? Robust Optimal Monetary Policy in a Forward-Looking Model." *Macroeconomic Dynamics*, 6, 111-144.
- Giannoni, M. and M. Woodford. 2002. "Optimal Interest Rate Rules: I. General Theory." *National Bureau of Economic Research Working Paper no. 9491*.
- Gilboa, I. and D. Schmeidler. 1989. "Maximin Expected Utility with Non-Unique Priors." *Journal of Mathematical Economics*, 18, 141-53.
- Gustafson P. and B. Clarke. 2004. "Decomposing Posterior Variance." *Journal of Statistical Planning and Inference*, 119, 311-327.
- Hansen, L. and T. Sargent. 2001a. "Acknowledging Misspecification in Macroeconomic Theory." *Review of Economic Dynamics*, 4, 519-35.
- Hansen, L. and T. Sargent. 2001b. "Robust Control and Model Uncertainty." Unpublished paper. Hoover Institution, Stanford University.
- Hansen, L. and T. Sargent. 2002. "'Certainty Equivalence' and 'Model Uncertainty'." Unpublished paper. Hoover Institution, Stanford University.
- Hansen, L. and T. Sargent. 2003. *Robust Control and Economic Model Uncertainty*. Book manuscript. Hoover Institution, Stanford University.
- von Hayek, F. 1942. "Scientism and the Study of Society." *Economica*, 9, 35, 267-91.

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

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

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

Levin, A., V. Wieland, and J. Williams, (1998), "Robustness of Simple Monetary Policy Rules Under Model Uncertainty." *National Bureau of Economic Research Working Paper no. 6570*.

Levin, A. and J. Williams. 2003. "Robust Monetary Policy with Competing Reference Models." *Journal of Monetary Economics*, 50, 945-975.

Marcellino, M. and M. Salmon. 2002. "Robust Decision Theory and the Lucas Critique." *Macroeconomic Dynamics*, 6, 1, 167-185.

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

Onatski, A. and N. Williams. 2003. "Modeling Model Uncertainty." *Journal of the European Economic Association*, 1, 1078-1122..

Pesaran, M. H., D. Pettenuzzo, and A. Timmermann, (2004), "Forecasting Time Series Subject to Structural Breaks: A Bayesian Regime Averaging Approach." Unpublished paper, University of Cambridge.

Raftery, A., D. Madigan, and J. Hoeting. 1997. "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association*, 92, 437, 179-91.

Rudebusch, G. 2002. "Assessing Nominal Income Rules for Monetary Policy with Model and Data Uncertainty." *Economic Journal*, 402-432.

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

Sims, C. 2002. "The Role of Models and Probabilities in the Monetary Policy Process." *Brookings Papers on Economic Activity*, 2, 1-40.

Taylor, J. 1993. "Discretion Versus Policy Rules in Practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195-214.

Tetlow, R. and P. von zur Muehlen. 2001. "Robust Monetary Policy With Misspecified Models: Does Model Uncertainty Always Call for Attenuated Policy?" *Journal of Economic Dynamics and Control*, 25, 6-7, 911-949.

von zur Muehlen., P.. 2001. "Activist vs. Non-Activist Monetary Policy: Optimal Rules Under Extreme Uncertainty (A Primer on Robust Control)." *Federal Reserve Board Finance and Economics Discussion Paper Series 2001-2*.

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

Woodford, M. 2003. "Optimal Interest Rate Smoothing." *Review of Economic Studies*, 70, 861-886.

Wright, J. 2003a. "Bayesian Model Averaging and Exchange Rate Forecasting." *Federal Reserve Board International Finance Discussion Papers 779*.

Wright, J. 2003b. "Forecasting US Inflation by Bayesian Model Averaging." *Federal Reserve Board International Finance Discussion Papers 780*.