Is the Original Taylor Rule Enough?
Simple versus Optimal Rules as Guides to Monetary Policy

William A. Brock, Steven N. Durlauf, James M. Nason, and Giacomo Rondina[1]

November 1, 2006

Conference Version: Preliminary and Incomplete

[1]University of Wisconsin, University of Wisconsin, Federal Reserve Bank of Atlanta, and University of Wisconsin respectively. The views herein represent those of the authors and are not those of either the Federal Reserve Bank of Atlanta, the Federal Reserve System, or any of its staff.

“..hold your tongue, I charge you, for the case is manifest…Did I not charge you not to speak? ‘twas plain enough before; and now you have puzzled it again.”

Gripus, the Theban Judge
***Amphitryon,* John Dryden**

1. Introduction

This paper contributes to the monetary policy evaluation literature by developing new strategies to study a few alternative policy rules that are the core of recent research. We are interested in assessing the case that can be made for optimal rules as opposed to simple rules for monetary policy. In our context, an optimal rule represents the solution to an intertemporal optimization problem in which a loss function for the policymaker and an explicit model of the macroeconomy are specified. By a simple rule, we refer to heuristics that embody the intuitions policymakers and economists have about how a central bank should react to aggregate disturbances. The classic example of a simple heuristic is Friedman’s *k*% money growth rule.

Modern versions of simple monetary policy rules typically embody a notion of “leaning against the wind”.[2] In this case, a central bank plays an active role in stabilizing macroeconomic aggregates. The canonical modern example of a simple policy heuristic is Taylor’s rule (1993), in which the Federal funds rate is increased in response to output relative to trend and by more than one-for-one with the inflation rate. Taylor’s rule is motivated by the notion that a central bank should attempt to reverse aggregate demand shocks, whereas aggregate supply shocks should be allowed to work unhindered in the economy.

The distinction between optimal and simple rules has taken on particular significance in recent debates over the desirability of inflation targeting, as exemplified in the competing views expressed in Svensson (2003) and McCallum and Nelson (2005). The latter paper takes particular interest in the claim made by Svensson that inflation targeting rules are preferable to instrument rules because they can embody the restrictions

[2]Tobin (1983) distinguishes between fixed and reactive rules. Modern work focuses on the latter.

intertemporal optimality places on the solution to the optimal policy problem.[3] As a theoretical matter, McCallum has shown that in principle an instrument rule may arbitrarily well approximate a targeting rule, so long as the same information set is available for conditioning. Nonetheless, Svensson's claim has force when targeting rules are compared to simple rules of the type in which we are interested.[4]

In this paper, we question the case for optimal rules at two levels. First, we consider the role model uncertainty plays in vitiating the superiority of optimal to simple rules. As argued by McCallum and Nelson (2005)[5], the optimality of a rule with respect to a given economic environment begs the question of its evaluation when the true model of the economy is not known. Our paper evaluates this issue by comparing the original Taylor rule with a model-specific optimal interest rule in the presence of various types of model uncertainty. Second, we compare simple and optimal rules from the perspective of what we call preference uncertainty. By preference uncertainty, we refer to uncertainty about the assumption that the objective function of a policymaker only depends on the variances of various states and controls. Our specific objective is to compare rules with respect to their effects on the variance of states at different frequencies. When a policymaker associates different losses to fluctuations at different frequencies, an analyst faces uncertainty as to which frequencies matter.

The analyst needs to study how rules drive fluctuations in various macroeconomic aggregates frequency by frequency to understand the role preference uncertainty plays in policymaking. Conventional policy assessments assume that the preferences of the policymaker are summarized by the variances of the aggregates under study. As we discuss below, such preferences ignore the disparate short-, medium- and long-run affects of policy. Since Brock and Durlauf (2004,2005) and Brock, Durlauf, and Rondina (2006) show that the choice of a policy rule yields a frequency by frequency variance tradeoff,

[3]Some of the arguments that appear in the literature strike us as self-evidently weak; for example, the claim that a targeting rule uses the entire information set of the policymaker. It is obvious that any information available for a targeting rule may be exploited by an instrument rule.

[4]Notice that the comparison of targeting rules to simple instrument rules renders moot the claim that targeting rules are more transparent than instrument rules.

[5]See McCallum (1988,1999) for early discussions of model uncertainty and monetary policy that broadly study many of the questions we wish to consider.

we argue that these tradeoffs should be reported to a policymaker, to account for different preferences with respect to fluctuations at different cycles. We wish to be clear that when we refer to preference uncertainty, we do not mean that the central bank does not know its own preferences. Rather, we believe it is important to report information on policy comparisons that avoid assuming a particular loss function for the central bank. For our purposes, this means that we wish to draw comparisons that reflect the possibility that the policymaker cares about fluctuations at some cycles more than others.

Our focus will be on monetary policy environments in which one engages in a binary comparison between a simple rule and a model-specific optimal alternative. Why focus on such stark comparisons between simple and optimal rules? We believe this exercise is useful in understanding how strong the case is for optimal approaches to monetary policy. The exercise matters because this is the decision problem a central bank must address when choosing one rule over another. We do not claim that our analysis proves either that the simple rules are superior to model-specific optimal rules in general or that the Taylor rule is superior to an inflation targeting rule in particular. Rather, we want to evaluate the strength of support in favor of optimal rules when compared to a standard rule of thumb for monetary policy; unlike Dryden's judge we think it important to understand when a case is not settled. Our conclusion is that there are not compelling reasons to favor an optimal rule over a rule of thumb, at least for the contexts we study.

As such, our analysis is in the spirit of Friedman (1948) whose conceptualization of a framework in which money growth is constant is very much couched in an explicit consideration of the information limits faced by a central bank. Friedman concludes his classic paper with (pg. 263)

> …I should like to emphasize the modest aim of the proposal. It does not claim to provide full employment…It does not claim to eliminate entirely cyclical fluctuations in output and employment. Its claim to serious consideration is that it provides a stable framework of fiscal and monetary action, that it largely eliminates the uncertainty…of discretionary actions by government authorities…It is not perhaps a proposal that one would consider at all optimum if our knowledge of the fundamental causes of cyclical fluctuations were considerably greater than I think them to be; it is a proposal that involves minimum reliance on uncertain and untested knowledge.

Model uncertainty has, due to the pioneering work of Hansen and Sargent ((2003) and especially (2006)) on robustness, become a major feature of current research.[6] In this paper we follow an approach developed in Levin and Williams (2003) and Brock, Durlauf, and West (2003, 2006) to evaluate policies in the presence of model uncertainty. The idea of this approach is to first construct a model space that includes all candidate models for the economy, evaluate policies for each of the candidate models, and then determine how to draw policy influences given the fact that the true model is unknown. This approach differs from the usual robustness analysis in that we do not focus on model misspecification which is measured local to a given baseline model.

Given a model space, it is necessary to determine how to aggregate information on rule performance across the different models. In this paper, we consider non-Bayesian approaches that are based on minimax and minimax regret criteria. These approaches involve guarding against "worst case scenarios", but do so in different ways that we formalize below. Relative to the Bayesian approach described below, neither minimax nor minimax regret requires the policymaker to take a stance on model space priors in order to compute posterior model probabilities.

Our analysis *does not* address the question of the design of optimal rules in the presence of model uncertainty. We regard this as an important, but distinct question from understanding the sensitivity of rules to model uncertainty. We believe our exercise is useful because it provides insights into how to evaluate the performance of rules in the presence of model uncertainty that cannot be explicitly described in advance. There are fundamental conceptual questions involved in thinking about policy effects when one cannot specify the support of the uncertainty faced by a policymaker. Nevertheless, we believe insights into this difficult problem are found from understanding how rules perform within a model space when, by construction, the rules fail to account for model uncertainty. Our approach to non-Bayesian decision theory follows Hansen and Sargent (who focus on minimax evaluation). Since we operate within somewhat different model spaces, it requires additional justification for our approach.

---

[6]Examples that focus on monetary policy include Giannoni (2002), Onatski and Stock (2002) and Tetlow and von zur Muehlen (2001).

We follow Levin in Williams (2003) in placing primary focus on model uncertainty with respect to forward- and backwards-looking elements to inflation. In the context of the Phillips curve, the 2005 special issue, "*The econometrics of the New Keynesian price equation*", in the *Journal of Monetary Economics* reveals little consensus about the response of inflation to forward- and backward-looking expectations. In that issue, the Rudd and Whelan paper is especially clear about how disagreements over inflation dynamics create uncertainty for new Keynesian model building.

Section 2 describes our basic framework for assessing model uncertainty. Section 3 discusses the evaluation of frequency-specific effects of policies. Section 4 analyzes some standard single input single output systems. Section 5 considers a standard two equation IS/PC model and focuses on comparing Taylor's original rule with a model- and preference-specific optimal alternative. Section 6 provides summary and discussion of future research directions.

2. Model uncertainty: basic ideas

At an abstract level, the issue of model uncertainty is easily described. Suppose that a policymaker is choosing from a set P of alternative policies. Given Θ, which captures whatever features of the economy matter to the policymaker, a policymaker experiences a loss associated with those features

$$l(\Theta). \tag{1}$$

The features defined by Θ are not, for interesting cases, a deterministic function of a policy and so are in general described by a conditional probability

$$\mu(\Theta | p, d, m) \tag{2}$$

where we explicitly allow the state of the economy to depend on the policy choice, p, the available data, *d*, and the model of the economy that is assumed to apply, *m*. Data dependence in this conditional probability may incorporate both direct dependence, as occurs in an autoregression, where forecasts depend on past realizations, or uncertainty about parameters. For simplicity, we measure losses as positive, so a policymaker wishes to minimize (1) by choosing a policy to affect (2). Our goal in the analysis is to relax the assumption that the policymaker knows the correct model of the economy, *m*. Rather, we assume that the policymaker knows that the correct model lies in a model space *M*. This space reflects uncertainty about the appropriate theoretical and functional form commitments needed to allow quantitative analysis of the effects of a policy.

Why treat model dependence explicitly in this formulation while subsuming other sources of uncertainty about Θ? Hansen and Sargent (2006) provide an extensive discussion of these and other issues; see also Brock, Durlauf, West (2003). We see two arguments as particularly important. One has to do with preferences: there are reasons to believe that individual preferences may treat model uncertainty differently from other types of uncertainty. These arguments are often based on experimental evidence of the type associated with the Ellsberg paradox, in which individuals appear to evaluate model uncertainty differently from other types. Put differently, the experimental evidence suggests that preferences exhibit ambiguity aversion as well as risk aversion, when model uncertainty is considered. For our purposes, what matters is that the evidence of ambiguity aversion suggests a tendency to guard against the worst case scenario with respect to models. A second justification concerns the levels at which model uncertainty occurs. As discussed in Brock, Durlauf, and West (2003,2006), model uncertainty exists at different levels: theory uncertainty, specification uncertainty, and parameter uncertainty. These different levels may have different salience; for example we do not think there is good evidence that parameter uncertainty within tightly specified models is as important a source of model uncertainty as differences in theoretical commitments to the presence of forward- and backward-looking elements in the Phillips curve. These differences are best accommodated by treating model uncertainty as distinct from data dependence.

The distinction between model uncertainty and other types of uncertainty is not unambiguous. For example, it is often the case that models may be nested, so that a given model may differ from another because a coefficient is zero. To be concrete, one can nest a finite model space which contains a pure backward-looking and pure forward-looking Phillips curve by considering a model space which contains a continuum of models which consist of weighted averages of the models. Any analysis of model uncertainty necessarily depends on prior judgments as to what models are of interest.

In this paper we focus on non-Bayesian decision criteria with respect to model uncertainty. In order to understand why we take this route, recall that the standard Bayesian solution to the optimal policy problem solves

$$\min_{p\in P} E\left(l\left(\Theta\right)\middle|p,d,M\right)=\sum_{m\in M} E\left(l\left(\Theta\right)\middle|p,d,m\right)\mu\left(m\middle|d\right) \tag{3}$$

where $\mu\left(m|d\right)$ denotes the posterior probability that model m is the correct one.[7,8] Recalling that

$$\mu\left(d|m\right)\propto\mu\left(m|d\right)\mu\left(m\right) \tag{4}$$

the Bayesian approach thus requires explicit assignment of model priors. The assignment of these priors is problematic since a researcher rarely possesses any meaningful prior knowledge from which priors may be constructed. The default prior in the model uncertainty literature assigns equal probabilities to all elements of the model space; but as discussed in Brock, Durlauf, and West (2003) such an approach is hard to justify. Hence, a first limitation of Bayesian approaches to model uncertainty is the requirement of introducing priors that have little meaning. There has been work on

[7]We do not address the issue of how to interpret these probabilities when none of the models is true. This is a deep and unresolved problem; see Key, Perrichi and Smith (1998) for some efforts to address.
[8]The Bayesian solution to model uncertainty involved model averaging, an idea suggested in Leamer (1978) and developed in detail in the work of Adrian Raftery, e.g. Raftery, Madigan, and Hoeting (1997).

evaluating the robustness of Bayesian inferences to prior choice in the statistics literature as well as efforts in the econometrics literature to develop alternatives to the diffuse priors that are used in the model uncertainty literature, but neither of these approaches is well understood for the context we study. We should note that the problem of defining priors continues to be a barrier to the adoption of Bayesian methods; Freedman (1991) is an example of a major statistician who has rejected Bayesianism for this reason.

Beyond the issue of priors, we feel that there are other good reasons, for at least some exercises, to avoid weighting models by posterior weights. These have to do with the interpretation of the likelihood component in (4), i.e. $\mu(m|d)$. In our judgment, the utility of models is context dependent. A model which fits the overall data well, i.e. receives a high value of $\mu(m|d)$, may perform relatively poorly when employed to consider regime changes when compared to another model whose overall fit is poorer. It is easy to see how this could be the case when the model which fits historical data less well is relatively immune to the Lucas critique.

Finally, it is important to recognize that the model spaces studied in contemporaneous macroeconomics have not arisen *sui generis*; they reflect a history of data work and as such, their relative fits reflect the evolution of empirical work so that one type of model may be evolved to achieve better fit than another because of data mining reasons. In this sense, one should not reify a model space as representing the unique set of candidate models that were available to a researcher prior to data analysis. This is not to say that model spaces do not reflect prior theoretical differences of course; rather that the econometric versions of models that are under consideration at a point in time have been influenced by the path of empirical work.

Non-Bayesian methods attempt to avoid dependence of inferences on model probabilities, typically by looking for extremes in behavior. At an intuitive level, these approaches abandon the search for optimal rules (in the Bayesian decision theory sense) in favor of good rules, where good is equated with the notion that the rule works relatively well regardless of which model is true. The minimax approach, proposed by Wald (1950), chooses a policy so that the loss under it is minimized under the least favorable model (relative to that rule) in the model set. Formally the minimax policy choice is

$$\min_{p\in P} \max_{m\in M} E\left(l(\Theta)\middle| p,d,m\right). \tag{5}$$

This formulation has been pioneered by Hansen and Sargent for macroeconomic contexts and forms the basis of the literature on robustness analysis. From the perspective of the economic theory literature, versions of minimax have been justified by Gilboa and Schmeidler (1989) among others. This type of work has in fact suggested that in addition to risk aversion (which leads to Bayesian expected loss calculations) agents can also exhibit ambiguity aversion, which applies to environments where probabilities cannot be assigned to possible outcomes (in our case models). The ambiguity aversion literature provides axiomatic justifications for minimax. Some of these, e.g. Epstein and Wang (1994), suggest that policies should be evaluated based on a weighted combination of losses under the least favorable model and the expected loss across models, an idea originally due to Hurwicz (1951); this is done in Brock, Durlauf, and West (2003), but we do not do so here given our desire not to assign model probabilities.

A standard criticism of the minimax criterion is that it is inappropriately conservative as it assumes the worst case possible in assessing policies. The force of this criticism, in our view, depends on whether the policy choice is driven by a relatively implausible model. This of course leads back to the question of model probabilities, the reasons for eschewing we have already given. As it turns out, there is an alternative non-Bayesian approach which moves away from the idea that the worst outcome should always be assumed: minimax regret. Minimax regret was originally proposed by Savage (1951) explicitly in order to avoid the pessimism of minimax. This approach focuses on the comparison of the effects of a given policy with the optimal one given knowledge of the true model. For a given policy *p* and model *m*, regret $R(p,d,m)$ is defined as

$$R(p,d,m) = E\left(l(\Theta)\middle| p,d,m\right) - \min_{p\in P} E\left(l(\Theta)\middle| p,d,m\right). \tag{6}$$

In words, the regret associated with a policy and model measures the loss incurred by the policy relative to what would have been incurred had the optimal policy been based upon the model. The minimax regret policy is correspondingly defined by

$$\min_{p\in P} \max_{m\in M} R\left(p,d,m\right). \tag{7}$$

When there are only two policies p_1 and p_2 under consideration, it is convenient to work with measures of maximum regret *MR*

$$MR\left(p_1\right) = \max_{m\in M} E\left(l\left(\Theta\right)\middle|p_1,d,m\right) - \min_{m\in M} E\left(l\left(\Theta\right)\middle|p_2,d,m\right) \tag{8}$$

and

$$MR\left(p_2\right) = \max_{m\in M} E\left(l\left(\Theta\right)\middle|p_2,d,m\right) - \min_{m\in M} E\left(l\left(\Theta\right)\middle|p_1,d,m\right). \tag{9}$$

Brock (2006) and Manski (2006) show that the minimax regret problem can be written as

$$\min_{\delta\in[0,1]} \left\{\delta MR\left(p_1\right), \left(1-\delta\right) MR\left(p_2\right)\right\}. \tag{10}$$

They further show that the optimal choice of δ is

$$\delta = \frac{MR\left(p_2\right)}{MR\left(p_1\right)+MR\left(p_2\right)}. \tag{11}$$

Notice that the weighting parameter δ, which will represent how one should weight the two candidate policies when making a policy choice, will depend smoothly on the maximum regret values; as such minimax regret avoids the property of minimax analysis that small changes in the payoffs of policies can lead to discontinous changes in the optimal policy choice.

Primarily due to seminal papers by Manski (2005,2006)[9], minimax regret has recently received a great deal of attention in the microeconometrics literature; but as far as we know it has not been used to study macroeconomic questions in general, let alone with respect to questions associated with policy evaluation.

One objection to the minimax regret criterion is that decisionmaking under it is not required to obey the axiom of independence of irrelevant alternatives (IIA), something first shown in Chernoff (1954). It is not clear that this axiom is appropriate for policy decisions, as opposed to individual ones. Without focusing on the individual/policymaker distinction, we can see reasons why the axiom may not be natural. The axiom in essence states that one's preferences over different pairs of actions does not depend on the context in which they are made, where the total choice set is part of the context. This is not an obvious requirement of rationality.[10] In absence of the axiom, there do exist coherent axiom systems that produce minimax regret as the standard for evaluation; Stoye (2006) is the state of the art treatment. Further, we would note that there are many results in the behavioral economics literature that speak against the axiom; indeed, as discussed in Camerer and Loewenstein (2003), what are known as context effects in behavioral economics directly challenge the IIA assumption.

We do not predicate our interest in minimax regret on a strong adherence to any of these arguments against the IIA assumption. For our purposes, we follow Manski and use minimax as a way of understanding policy effects without insisting on the superiority of minimax or any other decision rule.

A second issue that arises in our calculations concerns the role of randomized strategies as candidates for central bank policy. In our analysis, we will have recourse to allow for randomization between the Taylor rule and its model-specific optimal

[9]Chamberlain (2001) is an important precursor in suggesting minimax regret as an approach.

[10]Blume, Easley, and Halperin (2006) argue, for example, that an appropriate axiomatization of rationality should not require that agents have preferences which obey certain consistency conditions over all possible states of the world, as is required by Savage, but rather that rationality requires certain consistency conditions only over states under consideration. While they do not pursue this approach as a critique of irrelevance of independent alternatives, their work can be used to justify its exclusion as a rationality requirement.

counterpart. This is a corollary of our desire to treat the comparison between the two rules as a decision problem and to use minimax regret for assessment: minimax regret solutions can involve randomized strategies. Notice that to the extent that we conclude that the case for an optimal rule over original Taylor rule is weak i.e. that the weight placed on the Taylor rule is relatively large, it is *a fortiori* the case that the evidence would be even weaker if other rules were added to the choice set facing the policymaker.

In this application we treat a minimax regret policymaker as using the fractions suggested by the minimax regret calculation as weights to put on each rule, i.e. we posit here that a central bank simply uses the minimax regret calculation as an heuristic device to "hedge its bets" by choosing its policy to be the mix (suggested by the minimax regret calculation) of the two under consideration. Thus, our suggested implementation of minimax regret does *not* subject the public to the vagaries of the Federal Reserve flipping a biased coin and choosing interest rate policy at random at each FOMC meeting. It is of interest to ask what level welfare will actually be produced by using a mix of two policies with positive weights. If welfare is concave in the policies, then the mixed policy assures the policymaker that it will get at least as much welfare as a random policy. In our exercises for backwards looking models, the maximization objective is concave and the constraints are linear, so the welfare will indeed be concave in the policies. That said, we are using the minimax regret weights as measures of support for the two policies, not for randomization.

3. Frequency-specific effects of policies

As described above, we are also interested in understanding rule performance with respect to fluctuations at different frequencies. Standard analyses of the effects of policies focus on weighted averages of the unconditional variances of the states and possibly controls. Hence, it is common to compare policies by considering, for example, a weighted average of the unconditional variances of inflation and unemployment. Such calculations thus mask the effects of policies on the variance of fluctuations at different frequencies. Put differently, since the variance of a time series is the integral of its spectral density, analyses of the integral will mask the effects of a policy on fluctuations at different frequencies.

As a positive matter, these frequency-specific effects are of interest. As argued in Brock and Durlauf (2004,2005) and Brock, Durlauf, and Rondina (2006), there can exist fundamental tradeoffs in the variance associated with different frequencies of state variables. Brock, Durlauf, and Rondina (2006) provide a general formal treatment. For our purposes, it is sufficient to note that (abstracting away from issues of unit and explosive roots), one can always interpret the effect of a feedback control rule

$$u_t = \pi(L) x_{t-1} \tag{12}$$

on the bivariate dynamical system

$$A_0 x_t = \beta E_t x_{t+1} + A(L) x_t + B(L) u_t + W(L) \varepsilon_t \tag{13}$$

in the following way. Let x_t^{NC} denote the system when $u_t \equiv 0$, i.e. there is no control; $f_x^{NC}(\omega)$ is the associated spectral density matrix.[11] Each choice of control produces a sensitivity function $S^C(\omega)$ such that

$$f_x^C(\omega) = S^C(\omega) f_x^{NC}(\omega) S^C(\omega)'. \tag{14}$$

Different choices of the feedback rule thus shape the spectral density of the state variables relative to the no control baseline. Hence, one can characterize what feedback rules can achieve in terms of frequency-by-frequency effects on the state by characterizing the set of sensitivity functions that are available to the policymaker. This is the key idea underlying the notion of design limits to policy choice.

[11]When the uncontrolled system contains explosive or unit roots in the AR component, then the spectral density will not exist. Our formal results on restrictions on the sensitivity function still hold in this case. Our use of the spectral density of the uncontrolled system is done for heuristic purposes.

As shown in Brock, Durlauf, and Rondina (2006), for bivariate systems there is a simple way to characterize the set of feasible sensitivity functions. For each specification of a model of the uncontrolled system *M* and a control *C*, it must be the case that the associated sensitivity function obeys

$$\int_{-\pi}^{\pi} \log\left(\left|\det S\left(e^{-i\omega}\right)\right|^{2}\right) d\omega = K\left(M, C\right). \tag{15}$$

Brock, Durlauf, and Rondina show that whenever the system has no forward-looking elements, i.e. $\beta \equiv 0$, the constraint $K\left(M, C\right)$ will only depend on the unstable autoregressive roots of the unconstrained system; the constraint will equal 0 if there are no such roots, and is positive otherwise. This means that it is impossible to induce uniform reductions in the spectral density matrix of the no control system, i.e. reductions at some frequencies require increases at others. Of course, the overall variance may be reduced. For forward-looking systems, the results are more complicated. It may be possible to reduce the variance at all frequencies; nevertheless different rules still induce different tradeoffs. Brock, Durlauf and Rondina show that a variance minimizing rule may be one that increases the variance at some frequencies, even when a uniform reduction is possible.

As a normative matter, should a central bank evaluate rules on the basis of frequency specific effects? One reason, which is discussed in Onatski and Williams (2003) and which we pursue in Brock, Durlauf, Nason, and Rondina (2006) with respect to the effects on policy evaluation of alternative detrending methods, concerns the problem of measuring low frequency components of state variables. Low frequency behavior is of course hard to measure. Such measurement limitations take on particular importance if it is believed that long-run movements in the data are those outside the control of the central bank. For these reasons, it may seem reasonable for a policymaker to compare the behavior of rules over integrals of “trimmed spectral densities,” i.e. integrating spectral densities between $[-\pi,-\varsigma]\cup[\varsigma,\pi]$ rather than $[-\pi,\pi]$.[12]

A second argument for considering frequency specific effects is based on the nonseparable preferences. Under nonseparable preferences by a policymaker, different weights are assigned to different frequencies in computing expected welfare; Otrok (2001) has a nice example where the weights on different frequencies can differ by more than 9:1. An important insight of Otrok is that if preferences exhibit habit persistence, this means that the loss associated with volatility at high frequencies is greater than the loss associated with volatility at low frequencies. This type of reasoning suggests that it is important to consider how policies affect spectral densities over different intervals.

The upshot of this discussion is that we believe part of the communication exercise for policymakers should involve reporting frequency specific comparisons.

4. Model uncertainty and monetary policy evaluation: basic ideas

a. parameter uncertainty

In this section we illustrate our suggestions on how to compare simple and optimal policies in the context of a very stylized model consisting of a single state x_t described by

$$x_t = ax_{t-1} + u_t + \varepsilon_t \tag{16}$$

with associated control u_t which obeys a feedback rule of the form

$$u_t = -fx_{t-1}. \tag{17}$$

[12]Note that we are referring to integration of the spectral density in computing losses, not to the data per se. So this argument does not constitute a justification for the Hodrick-Prescott filter.

The policymaker's objective is to compare two candidate feedback rules, i.e. the policymaker can choose his actions from the action space $f \in \{a_1, a_2\}$. The model space is

$$M: \; a \in [\underline{a}, \overline{a}]. \tag{18}$$

We take the density μ_a to be symmetric on support $[\underline{a}, \overline{a}]$. We assume that

$$\underline{a} < a_1 < \frac{\underline{a} + \overline{a}}{2} < a_2 < \overline{a}. \tag{19}$$

Suppose that the objective function of the policymaker is to minimize the unconditional variance of the state variable; for each model the loss function is therefore $E\left(x_t^2 \mid f, m\right)$. Note that the control is costless in that its variance does not enter the loss function.

How do the different decision rules solve this problem? When policies are evaluated using expected losses, the policy problem is

$$\min_{f \in \{a_1, a_2\}} \int_{\underline{a}}^{\overline{a}} \frac{\sigma_\varepsilon^2}{1-(a-f)^2} d\mu_a = \min_{f \in \{a_1, a_2\}} \frac{\sigma_\varepsilon^2}{2}\left(\log\left(\frac{1+\overline{a}-f}{1-\overline{a}+f}\right) - \log\left(\frac{1+\underline{a}-f}{1-\underline{a}+f}\right)\right) \tag{20}$$

so long as neither of the policies produces a unit or explosive root in x_t.

The minimax problem for the policymaker is

$$\min_{f \in \{a_1, a_2\}} \max_{a \in [\underline{a}, \overline{a}]} \left(\frac{\sigma_\varepsilon^2}{1-(a-f)^2}\right) \tag{21}$$

which leads to the feedback rule

$$f = \begin{cases} a_1 & \text{if} \quad \underline{a} + \overline{a} < a_1 + a_2 \\ a_2 & \text{if} \quad \underline{a} + \overline{a} > a_1 + a_2 \end{cases} \tag{22}$$

independent of the relative plausibility of models in the model space. The behavior of the policymaker is controlled by the boundaries of the space; a finding discussed in a similar special case by Svensson (2002) and shown to hold for a general class of problems (when the model space is local) by Brock and Durlauf (2005).

Finally, to calculate the minimax regret solution, one first calculates the expression for maximum regret for each action, which are

$$\begin{gathered} MR(a_1) = \max_{a \in [\underline{a}, \overline{a}]} \left(\frac{1}{1-(a-a_1)^2} \right) - \min_{a \in [\underline{a}, \overline{a}]} \left(\frac{1}{1-(a-a_2)^2} \right) = \\ \frac{1}{1-(a_1-\overline{a})^2} - 1 = \frac{(a_1-\overline{a})^2}{1-(a_1-\overline{a})^2} \end{gathered} \tag{23}$$

and

$$\begin{gathered} MR(a_2) = \max_{a \in [\underline{a}, \overline{a}]} \left(\frac{1}{1-(a-a_2)^2} \right) - \min_{a \in [\underline{a}, \overline{a}]} \left(\frac{1}{1-(a-a_1)^2} \right) = \\ \frac{1}{1-(\underline{a}-a_2)^2} - 1 = \frac{(\underline{a}-a_2)^2}{1-(\underline{a}-a_2)^2}. \end{gathered} \tag{24}$$

In both cases the minimum loss under the alternative policy is always equal to the minimum loss reachable which is 1. The loss increases as the absolute distance between the action and the model $|a - a_i|$ is increased.

In contrast to the minimax case, under minimax regret the policymaker can optimally choose to play mixed strategies. Following the discussion in section 2, the optimal minimax regret rule computes $\min\left(\delta MR(a_1), (1-\delta) MR(a_2)\right)$ where

$$\delta = \frac{\left(\underline{a}-a_2\right)^2\left(1-\left(a_1-\overline{a}\right)^2\right)}{\left(\underline{a}-a_2\right)^2\left(1-\left(a_1-\overline{a}\right)^2\right)+\left(a_1-\overline{a}\right)^2\left(1-\left(\underline{a}-a_2\right)^2\right)}. \tag{25}$$

Notice that the magnitude of this probability is sensitive to all combinations of feedback parameters a_1 and a_2 and endpoints of the support of the autoregressive part of the state, $\underline{a}$ and $\overline{a}$. This is because, given a model, each rule is assessed according to how it performs versus its alternative. This leads to the important difference between an optimal policy under minimax regret compared to that under minimax: the continuity of δ with respect to the change in the model space parameters $\underline{a}$ and $\overline{a}$.

We can now relate this discussion to the comparison of a model specific optimal policy and a rule of thumb. Let a_1 be the rule of thumb parameter and $a_2 = \frac{\underline{a}+\overline{a}}{2}$ so that it represents the optimal feedback rule for (16).[13] Our notion of the rule of thumb is that the policymaker has some baseline feedback parameter that it wishes to consider as opposed to the optimal feedback rule. As such the rule of thumb is in the spirit of the original Taylor rule.

The minimax evaluation of the two rules yields a feedback parameter choice of the form

$$f = \begin{cases} a_1 & \text{if} \quad \frac{\underline{a}+\overline{a}}{2} < a_1 \\ \frac{\underline{a}+\overline{a}}{2} & \text{if} \quad \frac{\underline{a}+\overline{a}}{2} > a_1 \end{cases} \tag{26}$$

How do we interpret this? In essence, the rule of thumb is preferable to the optimal rule when there is danger that the optimal rule will overshoot too much.

The minimax regret approach yields, (see Technical Appendix for derivation) a weight for the rule of thumb of

[13]This is the optimal rule given our assumption of symmetry of μ_a.

$$\delta = \begin{cases} \dfrac{\left(\underline{a}-\overline{a}\right)^2\left(1-\left(\underline{a}-a_1\right)^2\right)}{\left(\underline{a}-a_1\right)^2\left(4-\left(\underline{a}-\overline{a}\right)^2\right)+\left(\underline{a}-\overline{a}\right)^2\left(1-\left(\underline{a}-a_1\right)^2\right)} & \text{if} \quad \dfrac{\underline{a}+\overline{a}}{2} < a_1 \\ \dfrac{\left(\underline{a}-\overline{a}\right)^2\left(1-\left(\overline{a}-a_1\right)^2\right)}{\left(\overline{a}-a_1\right)^2\left(4-\left(\underline{a}-\overline{a}\right)^2\right)+\left(\underline{a}-\overline{a}\right)^2\left(1-\left(\overline{a}-a_1\right)^2\right)} & \text{if} \quad \dfrac{\underline{a}+\overline{a}}{2} > a_1 \end{cases} \tag{27}$$

This formulation indicates how using minimax regret, the strength of the case for the optimal policy will depend on the extreme support points $\underline{a}$ and $\overline{a}$. When the rule of thumb is less aggressive than the optimal policy, its weight decreases the farther away it is from $\overline{a}$. So, in this sense the minimax regret approach penalizes the rule of thumb in a way that minimax does not.

We next consider the effects of policies at different frequencies. Let the loss function of the policymaker be such that he only cares about business cycle frequencies. We measure their role in fluctuations by focusing on the spectral density for $\omega \in BC = \left[-\frac{\pi}{2}, -\hat{\omega}\right] \cup \left[\hat{\omega}, \frac{\pi}{2}\right]$; for simplicity we choose $\hat{\omega}$ so that $\tan\left(\frac{\hat{\omega}}{2}\right) = \frac{1}{2}$. The loss function is thus $\int_{\omega \in BC} f_x(\omega) d\omega$. Let $\rho = a - f$. Then it is shown in the Technical Appendix that for any feedback rule f, the associated loss function is

$$\int_{\omega_L}^{\omega_H} \frac{1}{1+\rho^2 - 2\rho\cos\omega} d\omega = \frac{2}{1-\rho^2}\left(\arctan\left(\frac{1+\rho}{1-\rho}\right) - \arctan\left(\frac{1+\rho}{2(1-\rho)}\right)\right) \tag{28}$$

This expression reaches a maximum in the interval $\rho \in (-1,1)$. Its value as $\rho \to -1$ is lower than the value as $\rho \to 1$ for $\rho > 0$; a generic figure of the spectral density appears in the Technical Appendix. For the two candidate policies, the losses are

$$L_{BC}\left(\frac{\underline{a}+\overline{a}}{2},a\right)=$$

$$\frac{1}{\pi}\frac{\sigma_{\varepsilon}^{2}}{1-\left(a-\frac{\underline{a}+\overline{a}}{2}\right)^{2}}\left(\arctan\left(\frac{1+\left(a-\frac{\underline{a}+\overline{a}}{2}\right)}{1-\left(a-\frac{\underline{a}+\overline{a}}{2}\right)}\right)-\arctan\left(\frac{1+\left(a-\frac{\underline{a}+\overline{a}}{2}\right)}{2\left(1-\left(a-\frac{\underline{a}+\overline{a}}{2}\right)\right)}\right)\right) \quad (29)$$

and

$$L_{BC}\left(a_1,a\right)=$$

$$\frac{1}{\pi}\frac{\sigma_{\varepsilon}^{2}}{1-\left(a-a_1\right)^{2}}\left(\arctan\left(\frac{1+\left(a-a_1\right)}{1-\left(a-a_1\right)}\right)-\arctan\left(\frac{1+\left(a-a_1\right)}{2\left(1-\left(a-a_1\right)\right)}\right)\right). \quad (30)$$

respectively.

In each case, the loss function reaches a maximum and a minimum at autoregressive coefficients different from those that achieve the maximum and the minimum when integration is taken over all frequencies. The generic shape of the frequency specific loss is highest for a slightly positive autoregressive coefficient, denote this point as a_{BC}. The loss steadily reduces as the autoregressive coefficient is increased or decreased, however it tends to be lower for negative autoregressive coefficients. These loss functions lead to the minimax and minimax regret solutions.

The minimax policy associated with a given business cycle frequency loss function is

$$f=\begin{cases} a_1 \text{ if } \frac{\underline{a}-\overline{a}}{2}<a_{BC}<\frac{\overline{a}-\underline{a}}{2} \text{ but } \underline{a}-a_1>a_{BC} \text{ or } \overline{a}-a_1<a_{BC} \\ \frac{\underline{a}+\overline{a}}{2} \text{ if } \underline{a}-a_1<a_{BC}<\overline{a}-a_1 \text{ but } \frac{\underline{a}-\overline{a}}{2}>a_{BC} \text{ or } \frac{\overline{a}-\underline{a}}{2}>a_{BC} \\ a_1 \quad \frac{\underline{a}+\overline{a}}{2} \text{ indifferent otherwise} \end{cases} \quad (31)$$

We compare the minimax policy to the minimax regret under the assumptions $\frac{\underline{a}-\overline{a}}{2} < a_{BC} < \frac{\overline{a}-\underline{a}}{2}$ and $\underline{a}-a_1 < a_{BC} < \overline{a}-a_1$ so that from a minimax perspective the two policies are equally good. This happens because the maximum frequency-specific loss under both policies corresponds to the maximum loss reached at a_{BC}, which we denote by L_{BC}^*. Under this assumption, maximum regret under the two policies considered here depends on the best possible outcome under the alternative policy and this depends on the relative position of the two policies. To see this consider that the maximum regrets under the two policies are

$$M\left(\frac{\underline{a}+\overline{a}}{2}\right) = L_{BC}^* - \min_a L_{BC}\left(a_1, a\right) \tag{32}$$

and

$$M\left(a_1\right) = L_{BC}^* - \min_a L_{BC}\left(\frac{\underline{a}+\overline{a}}{2}, a\right) \tag{33}$$

Given the postulated functional form for the frequency specific variance over the model space, the first regret is bigger than the second when $a_1 > \frac{\underline{a}+\overline{a}}{2}$ while it is smaller in the opposite case. This implies the weights associated with the feedback parameter are

$$\delta\left(a_1\right) > .5 \text{ if } a_1 > \frac{\overline{a}+\underline{a}}{2}; \;\; \delta\left(a_1\right) < .5 \text{ if } a_1 < \frac{\overline{a}+\underline{a}}{2} \tag{34}$$

Given the conditions on the parameters assumed, while under the minimax approach any policy is equally good, under minimax regret the optimal weights are uniquely identified.

To see how these formulas translate into policy comparison, we consider the case where $\underline{a} = .05$, $\overline{a} = .95$, and the rule of thumb feedback is set at $a_1 = .3$ and $a_1 = .8$ respectively. Table 1 Panel A reports the losses associated with the optimal rule and each rule of thumb, where losses are first equated with the total variance of the state and equated with the variance associated with frequencies in the intervals $\left[0,\frac{\pi}{4}\right],\left[\frac{\pi}{4},\frac{\pi}{2}\right],\left[\frac{\pi}{2},\frac{3\pi}{4}\right]$, and $\left[\frac{3\pi}{4},\pi\right]$. The main features we would note are the following. First, from the perspective of total variance, the differences between the average loss under the optimal rule and rule of thumb are relatively small; the loss under the rules of thumb is less than 17% larger than the loss under the optimal rule. More interesting features are uncovered in the frequency specific analyses. For low frequencies $\left[0,\frac{\pi}{4}\right]$ and middle-to-low frequencies $\left[\frac{\pi}{4},\frac{\pi}{2}\right]$ whereas the rule of thumb $a_1=.8$ outperforms the variance minimizing rule as well. For middle-to-high frequencies $\left[\frac{\pi}{2},\frac{3\pi}{4}\right]$ and high frequencies $\left[\frac{3\pi}{4},\pi\right]$ the rule of thumb $a_1=.3$ outperforms the variance minimizing rule. This illustrates that the comparison of rules is sensitive to how one evaluates fluctuations of different cycles.

The translation of these different losses into minimax and minimax regret comparisons of rules is done in Table 1 Panel B. As suggested by the losses, while the minimax rule selects the variance minimizing rule over the two rules of thumb when overall variance defines the loss function, one chooses the rule of thumb $a_1=.8$ over the optimal rule for the intervals $\left[0,\frac{\pi}{4}\right]$ and $\left[\frac{\pi}{4},\frac{\pi}{2}\right]$;[14] the rule of thumb $a_1=.3$ is selected over the optimal rule for the intervals $\left[\frac{\pi}{2},\frac{3\pi}{4}\right]$ and $\left[\frac{3\pi}{4},\pi\right]$. Interestingly, the minimax regret weights tell a similar story. The optimal rule receives a large preponderance of the weight when total variance is considered, although the weight is always less than 6/7. For the frequencies intervals, the rule $a_1=.8$ receives more than half weighting for the lower frequencies and $a_1=.3$ more than half weighting for the higher ones.

Together, these findings illustrate how the comparison of rules is sensitive to how one evaluates fluctuations of different cycles. They also illustrate how using minimax regret gives a more nuanced view of the superiority of optimal rules. Even for the case

[14]Notice that this case perfectly matches the formula in (34).

where all frequencies are equally weighted in variance losses, some weight is warranted for the rules of thumb we have chosen. While we do not claim, of course, that this model calibrates to an actual macroeconomic stabilization context, we believe these results are suggestive.

b. theory uncertainty: backwards- versus forward- looking models

We now consider the evaluation of policies when model uncertainty concerns the role of backwards-looking versus forward-looking elements in determining aggregate outcomes. Here we work with an intertemporal loss function

$$\min_{\{x_t\}\{u_t\}} E_0\left[\sum_{t=0}^{\infty}\delta^t \frac{1}{2} L_t\right] \tag{35}$$

where $L_t = x_t^2$.

For a backwards-looking model we consider

$$x_t = a x_{t-1} + b_B u_{t-1} + \varepsilon_t;\ \varepsilon_t \text{ white noise.} \tag{36}$$

The optimal feedback rule for this model O_B is shown in the Technical Appendix[15] to be

$$u_t^B = -\frac{a}{b_B} x_t. \tag{37}$$

This is the same as the optimal feedback rule derived directly from the unconditional variance. Notice that when the disturbance is not white noise this rule is different and it will contain some further lagged terms of the state.

Our alternative forward-looking model is

[15] The optimal policy calculations performed in this section and Section 4.c are similar to those found in Clarida, Gali, and Gertler (1999), especially for forward looking models.

$$x_t = \beta E_t\left(x_{t+1}\right) + b_F u_{t-1} + \varepsilon_t;\ \varepsilon_t \text{ white noise.} \tag{38}$$

The optimal feedback rule (O_F) is shown in the Technical Appendix to be defined by

$$u_t^F = -\frac{\beta}{\delta b_F} x_t + \frac{\beta}{\delta} u_{t-1}^F . \tag{39}$$

The substantial smoothing typical of timeless perspective optimal rules in forward looking models is easy to see in this expression. An explicit expression for u_t in terms of current and past x_t can be derived if $\beta < \delta$.

In addition to O_F and O_B we consider two additional rules. The first is a rule of thumb (RoT)

$$u_t = -1.25 x_t \tag{40}$$

where our choice of the coefficient -1.25 is meant to capture the basic "Taylor principle" of reacting to innovations in the inflation series more than one for one. Second, we consider a restricted optimal rule (RO_F) for the forward-looking model, by which we mean rules restricted to the form $u_t = f x_t$ in which the parameter *f* is chosen to minimize (35). The restricted optimal rule is shown in Brock, Durlauf and Rondina (2006) to be

$$u_t^{RO} = \frac{\left(1 - \sqrt{1 + 8\beta^2} - 4\beta^2\right)}{b_F 8\beta^2} x_t . \tag{41}$$

We now consider how these different rules perform given uncertainty about whether the true model is backwards-looking or forward-looking. We focus on the parametrizations $\frac{\beta}{\delta} = \hat{\beta} = 0.9$, $a = 0.9$, and $\frac{b_B}{b_F} \in \{.1, 1\}$. We focus on these two values of $\frac{b_B}{b_F}$ as .1

appears empirically salient given previous research (see Rudebusch and Svensson (1999) and Rudebusch (2002), whereas 1 captures the case where forwards and backwards components produce equal effects and seems to us a more interesting theoretical benchmark. The left hand side panels in Figures 1 and 2 compare the spectral density produced, for the backwards model, by the optimal backwards rule with that produced by the optimal forwards rule, restricted optimal forward-looking rule, and the rule of thumb. The right hand side panels compare the spectral density produced, for the forward-looking model, by the optimal forward-looking rule with the other three rules. As before, in order to summarize the differences in the spectral densities, Tables 2 and 3 report the losses associated with each rule and each model, measured as the overall variance of the state and the variance of the frequencies for intervals $\left[0,\frac{\pi}{4}\right],\left[\frac{\pi}{4},\frac{\pi}{2}\right],\left[\frac{\pi}{2},\frac{3\pi}{4}\right],\left[\frac{3\pi}{4},\pi\right]$.

The first result of interest in Table 2 Panel A concerns the effects of the rules on overall variance when $\frac{b_B}{b_F}=.1$. When the true model is backwards looking, O_F, RO_F and RoT perform poorly when compared to the optimal rule for the model O_B. When the overall variances are decomposed into frequency interval contributions, it turns out that almost all of the difference is due to the low frequencies, i.e. cycles of period 8 or greater. The results are quite different when $\frac{b_B}{b_F}=1$ as indicated in Table 3 Panel A. In this case, O_F continues to perform quite poorly, but both RO_F and RoT perform well, close to the performance achieved by O_B. The associated decomposition indicates that poor performance of O_F is due primarily to its poor performance for periods of 4 to 8. The spectral densities of the state variable are reported in the left hand side panels in Figures 1 and 2.

The spectral densities clarify that the application of the optimal forward looking rule to the backwards looking model can induce cycles that do not appear under the other rules. To understand why this is so, observe that application of O_F to the backwards looking model yields

$$x_t = ax_{t-1} + b_B u^*_{t-1} + \varepsilon_t = ax_{t-1} - \frac{b_B \hat{\beta}}{b_F} \frac{1}{\left(1-\hat{\beta}L\right)} x_{t-1} + \varepsilon_t. \tag{42}$$

Solving this equation

$$\begin{gathered}\left(1-\hat{\beta}L\right)x_t = a\left(1-\hat{\beta}L\right)x_{t-1} - \frac{b_B}{b_F}\hat{\beta}x_{t-1} + \varepsilon_t\left(1-\hat{\beta}L\right) \Rightarrow \\ x_t = \left(\hat{\beta} + a - \frac{b_B}{b_F}\hat{\beta}\right)x_{t-1} - a\hat{\beta}x_{t-2} + \varepsilon_t\left(1-\hat{\beta}L\right).\end{gathered} \tag{43}$$

which produces an ARMA(2,1) representation for the state variable of the form

$$x_t = \frac{\left(1-\hat{\beta}L\right)}{1-\left(\hat{\beta} + a - \frac{b_B}{b_F}\hat{\beta}\right)L + a\hat{\beta}L^2}\varepsilon_t . \tag{44}$$

The roots of the polynomial in the denominator are

$$\frac{\left(\hat{\beta} + a - \frac{b_B}{b_F}\hat{\beta}\right) \pm \sqrt{\left(\hat{\beta} + a - \frac{b_B}{b_F}\hat{\beta}\right)^2 - 4a\hat{\beta}}}{a\hat{\beta}} . \tag{45}$$

It is easy to see that the discriminant $\left(\hat{\beta} + a - \frac{b_B}{b_F}\hat{\beta}\right)^2 - 4a\hat{\beta}$ can be negative for a wide range of values which suggests that the roots are complex conjugates. This happens for example under the parameterizations we have used. For such cases, the spectral density of the state produces a peak at business cycle frequencies (when the time series is quarterly). In contrast, the rule of thumb produces an AR(1) process

$$x_t = \frac{1}{1-\left(a+\frac{b_B}{b_F}\frac{\left(1-\sqrt{1+8\beta^2}-4\beta^2\right)}{8\beta^2}\right)L}\varepsilon_t . \qquad (46)$$

These general differences are reflected in the parameterized examples we have described.

Table 2 Panel B and Table 3 Panel B report analogous results when the forward looking model is the true one. One important finding is that the optimal backwards rule produces nonstationarity when $\frac{b_B}{b_F} = .1$. This leads to an infinite loss and an ill-defined spectral density. Interestingly the rule of thumb produces a stationary state variable; intuitively the rule of thumb avoids overreaction which occurs under the optimal backwards rule. When one considers the behavior of the frequency intervals, it is evident that the superiority of the optimal rule is largely driven by high frequency behavior; that said, the optimal rule outperforms the others in all the frequency intervals. .

When $\frac{b_B}{b_F} = 1$, the fact that all rules produce stationary state variables makes for more interesting comparisons. Relative to the optimal rule, the losses from these rules, as well as the rule of thumb, are substantially smaller than occurs when the forwards rules are applied to the backwards looking model. This suggests an important asymmetry in the effects of model uncertainty. When one considers the frequency decompositions, the sources of the poorer performance of the forward rules and rule of thumb are different from the forward-looking case. Here, the main source of inferior performance is the high frequencies.

How do these findings translate into the strength of the case for the model-specific optimal rules? Here we consider comparisons with the optimal forward rule. For $\frac{b_B}{b_F} = .1$, Table 2 Panel C indicates that under the minimax rule, the O_F is selected against O_B and RO_F when the loss function is defined by total variance, but is rejected in favor of RoT. When one restricts variances to intervals, then O_F is selected against any of the other rules for the intervals $\left[\frac{\pi}{2},\frac{3\pi}{4}\right]$ and $\left[\frac{3\pi}{4},\pi\right]$, reflecting its good performance for high frequencies under the backwards model. In contrast, RoT is selected for the intervals $\left[0,\frac{\pi}{4}\right]$ and $\left[\frac{\pi}{4},\frac{\pi}{2}\right]$ and RO_F is selected for $\left[\frac{\pi}{4},\frac{\pi}{2}\right]$. The rule O_B is never selected because the nonstationarity it induces in the forward-looking case. Turning to minimax regret, we see that the weight assigned to O_F versus either RO_F or RoT is slightly above ½. When one focuses on the frequency intervals, we see that the weight assigned to O_F depends on the interval, with a weight of 1 assigned to the optimal forward rule in the interval $\left[\frac{3\pi}{4},\pi\right]$ and qualitatively similar results in the interval $\left[\frac{\pi}{2},\frac{3\pi}{4}\right]$.

For the case $\frac{b_B}{b_F}=1$, the conclusions are rather different. As indicated in Table 3 Panel C, by the minimax criterion, the optimal forward rule O_F is never chosen when overall variance is considered. On the other hand, O_F is chosen versus the other rules when one focuses on $\left[\frac{3\pi}{4},\pi\right]$. Interestingly, the minimax regret analysis place very high weights on the alternatives to O_F for both overall variance and $\left[\frac{\pi}{4},\frac{\pi}{2}\right]$.

In terms of conclusions, one message we believe these exercises give is that the case for a particular rule is sensitive to model uncertainty and the role of frequency-specific fluctuations in overall losses. Overall, it is reasonable to say that the optimal rule O_F performed relatively well in the sense that it never induced instabilities so that in this sense it outperformed the other optimal rule O_B. But in comparing these rules, other comparisons depended critically on how one evaluates different frequencies and on the relative strength of the backwards and forward feedback elements for the economy. The case for O_F very much depends on a loss function that focuses on high frequencies.

***c*. hybrid models**

We now consider a hybrid model which allows for forward and backwards-looking elements.

$$x_t = (1-\theta)\beta E_t x_{t+1} + \theta a x_{t-1} + b(\theta) u_{t-1} + \varepsilon_t; \qquad \varepsilon_t \text{ white noise} \tag{47}$$

where $b(\theta) = \theta b_B + (1-\theta) b_F$. It can be shown that the fully optimal policy for this model under the objective function specified above is implemented by the instrument rule

$$u_t^*(\theta) = -\frac{(1-\theta)\beta + \theta a\delta}{b(\theta)\delta} \frac{\left(1 - \frac{\theta(1-\theta)a\beta}{(1-\theta)\beta + \theta a\delta} L\right)}{\left(1 - (1-\theta)\frac{\beta}{\delta} L\right)} x_t . \tag{48}$$

Notice that as $\theta \to 0$ this policy converges to the policy that is optimal under the purely forward looking model. The rational expectations equilibrium implemented by this rule is unique when

$$\frac{(1-\theta)\beta^2 + \delta}{\beta\delta} > 1. \tag{49}$$

The process for the state variable under this policy is

$$x_t = \frac{\delta}{(1-\theta)\beta^2 + \delta}\left(1 - (1-\theta)\frac{\beta}{\delta} L\right)\varepsilon_t . \tag{50}$$

Suppose now that the benchmark model is one with $\theta = \hat{\theta}$, however the model space ranges in $\theta \in [0,1]$. The instrument rule followed by the policy maker is thus

$$u_t^*\left(\hat{\theta}\right)=\frac{\alpha\left(\hat{\theta}\right)}{b\left(\hat{\theta}\right)}\frac{\left(1-B\left(\hat{\theta}\right)L\right)}{\left(1-\gamma\left(\hat{\theta}\right)L\right)}x_t \tag{51}$$

with associated terms $\alpha\left(\hat{\theta}\right)=-\frac{\left(1-\hat{\theta}\right)\beta+\hat{\theta}a\delta}{\delta}$, $B\left(\hat{\theta}\right)=\frac{\hat{\theta}\left(1-\hat{\theta}\right)a\beta}{\left(1-\hat{\theta}\right)\beta+\hat{\theta}a\delta}$, and $\gamma\left(\hat{\theta}\right)=\left(1-\hat{\theta}\right)\frac{\beta}{\delta}$. Again, we are interested in evaluating how this policy performs across the model space compared to a generic rule of thumb $\bar{u}_t=\rho x_t$ for $\rho<-1$.

The respective laws of motion for the state variable under each of the two policies are

$$x_t=\left(1-\theta\right)\beta E_t x_{t+1}+\theta a x_{t-1}+b\left(\theta\right)\rho x_{t-1}+\varepsilon_t \tag{52}$$

and

$$x_t=\left(1-\theta\right)\beta E_t x_{t+1}+\theta a x_{t-1}+b\left(\theta\right)u_{t-1}^*\left(\hat{\theta}\right)+\varepsilon_t\,. \tag{53}$$

Consider first the case of the rule of thumb. The solution for the state variable when it exists and is unique always takes the form of an AR(1) process:

$$x_t=\frac{1}{\lambda_1\left(\theta,\rho\right)\beta\left(1-\theta\right)}\frac{1}{\left(1-\lambda_2\left(\theta,\rho\right)L\right)}\varepsilon_t \tag{54}$$

where

$$\frac{1}{\lambda_i\left(\theta,\rho\right)}=\frac{1\pm\sqrt{1-4\left(1-\theta\right)\beta\left(\theta a+\rho b\left(\theta\right)\right)}}{2\left(\theta a+\rho b\left(\theta\right)\right)}\,. \tag{55}$$

For uniqueness and existence it has been assumed that

$$\left|\lambda_1(\theta,\rho)\right|>1,\left|\lambda_2(\theta,\rho)\right|<1. \tag{56}$$

This makes clear that a good rule of thumb is one where $\rho>-1$ because it ensures uniqueness and existence of a bounded solution, i.e. it ensures (54) holds under many specifications for θ; as such this is a version of the Taylor principle.

For the optimal rule in our generic hybrid model, it can be shown that the law of motion for the system is

$$x_t=(1-\theta)\beta\left(1-\gamma\left(\hat{\theta}\right)L\right)E_t x_{t+1}+H\left(\theta,\hat{\theta}\right)x_{t-1}+K\left(\theta,\hat{\theta}\right)x_{t-2}+\left(1-\gamma\left(\hat{\theta}\right)L\right)\varepsilon_t. \tag{57}$$

Where $H\left(\theta,\hat{\theta}\right)\equiv\gamma\left(\hat{\theta}\right)+\theta a+b(\theta)\frac{\alpha\left(\hat{\theta}\right)}{b\left(\hat{\theta}\right)}$ and $K\left(\theta,\hat{\theta}\right)=-\theta a\gamma\left(\hat{\theta}\right)-b(\theta)\frac{\alpha\left(\hat{\theta}\right)B\left(\hat{\theta}\right)}{b\left(\hat{\theta}\right)}$. The primary issue for evaluating this rule under model uncertainty is the following. When the policymaker is working with the correct model, the backwards-looking terms can be shown to be zero, i.e. $H(\theta,\theta)=0$ and $K(\theta,\theta)=0$, but when the model is not correct these terms are in general different than zero. This means that the dynamics of the system become non trivial. It can be shown that the moving average representation of the state variable, $x_t=G(L)\varepsilon_t$, is defined by

$$G(L)=\frac{\left(1-\gamma\left(\hat{\theta}\right)L\right)\left(G_0(1-\theta)\beta-L\right)}{\left((1-\theta)\beta-\left(1+(1-\theta)\beta\gamma\left(\hat{\theta}\right)\right)L+H\left(\theta,\hat{\theta}\right)L^2+K\left(\theta,\hat{\theta}\right)L^3\right)}. \tag{58}$$

This representation is neither unique nor bounded unless restrictions are imposed in the roots of the autoregressive polynomial. Writing the polynomial as

$$(1-\theta)\beta\left(1-\lambda_1\left(\theta,\hat{\theta}\right)L\right)\left(1-\lambda_2\left(\theta,\hat{\theta}\right)L\right)\left(1-\lambda_3\left(\theta,\hat{\theta}\right)L\right) \tag{59}$$

the condition for a unique and bounded solution is

$$\left|\lambda_1\left(\theta,\hat{\theta}\right)\right|>1,\left|\lambda_2\left(\theta,\hat{\theta}\right)\right|<1,\left|\lambda_3\left(\theta,\hat{\theta}\right)\right|<1. \tag{60}$$

We assume that this condition holds so that the solution is again an ARMA(2,1)

$$x_t=\frac{1}{\lambda_1\left(\theta,\hat{\theta}\right)(1-\theta)\beta}\frac{\left(1-\gamma\left(\hat{\theta}\right)L\right)}{\left(1-\lambda_2\left(\theta,\hat{\theta}\right)L\right)\left(1-\lambda_3\left(\theta,\hat{\theta}\right)L\right)}\varepsilon_t. \tag{61}$$

This again illustrates how optimal rules can be the source of fluctuations because of model uncertainty.

To make the analysis concrete, we perform a numerical calculation for the parameters $\beta=0.9$, $\delta=0.99$, $a=0.9$ As before, we consider $\frac{b_B}{b_F}\in\{.1,1\}$. We construct the optimal rule (OB) based on the benchmark model $\theta=.5$; this produces optimal rules $u_t^*(\theta)=-1.63\frac{(1-0.22L)}{(1-0.45L)}x_t$ for $\frac{b_B}{b_F}=.1$ for $u_t^*(\theta)=-0.90\frac{(1-0.22L)}{(1-0.45L)}x_t$ for $\frac{b_B}{b_F}=1$. As before, our rule of thumb is $\bar{u}_t=-1.25x_{t-1}$. In all cases, a unique bounded rational expectations equilibrium for the state variable exists. Figure 3 provides the overall and frequency interval performances of the two rules as well as the associated values of the Bode integral constraint. Tables 4 and 5 describe the minimum, maximum, and average losses for the rules as well as minimax and minimax regret calculations.

Figure 3 describes the relative performances of the optimal rule for the benchmark model versus the rule of thumb across the model space, for the case $\frac{b_B}{b_F}=.1$. As the Figure indicates, for overall variance the rule of thumb outperforms the benchmark optimal rule for smaller values of θ. This indicates an asymmetry in the effects of overestimating versus underestimating the role of forward-looking elements.

Underestimation leaves the *OB* superior to *RoT* whereas overestimation may not. When one considers the performance of the two rules for frequency intervals, we find that *RoT* uniformly outperforms *OB* for the high frequency interval $\left[\frac{3\pi}{4},\pi\right]$ whereas *OB* uniformly outperforms *RoT* for the low frequency interval $\left[0,\frac{\pi}{4}\right]$. Interestingly, the Bode constraint K is uniformly smaller for the rule of thumb, even at the benchmark value $\theta = .5$. This suggests that while *RoT* does not efficiently shape the spectral density of the state in order to minimize overall variance, by generating a weaker constraint, this inefficiency is partially overcome. We conjecture that an aspect of robustness that warrants further study is how different rules affect the Bode constraint.

Table 4 translates these findings into minimax and minimax regret comparisons. We find that minimax will choose *OB* when losses are measured by overall variance and by variance in the low frequencies. For the other frequency intervals, *RoT* is chosen. The minimax regret calculations indicate that for overall variance and for all 4 of the frequency intervals, substantial (greater than 40%) weight is assigned to the rule of thumb.

Figure 4 describes analogous results for $\frac{b_B}{b_F} = 1$. For this case, with respect to overall variance, *OB* uniformly outperforms *RoT* across the model space. In terms of frequency intervals, unlike the case where $\frac{b_B}{b_F} = .1$, *OB* strongly outperforms *RoT* at the high frequencies; for the other intervals *RoT* nearly always outperforms *RoT*. These differences lead to different minimax and minimax regret calculations. Under minimax, *OB* is chosen for overall variance and for high frequencies, whereas *RoT* is chosen for the other intervals. For minimax regret, substantial (at least 30%) weight is assigned to *RoT* except for the high frequencies, where a weight of 1 is assigned to *OB.*

To further illustrate how considerations of overall variance can mask important frequency-specific differences in policy comparisons, Figures 5 and 6 reports the spectral densities for the state that are produced by the two rules under three different values of θ

for the true model. Figure 5 assumes $\frac{b_B}{b_F} = .1$; Figure 6 assumes $\frac{b_B}{b_F} = 1$. First consider the case where the benchmark model is the true one, i.e. $\theta = .5$, so that the model-specific optimal rule is by construction variance minimizing. The figures indicate how sensitive the comparison of rules is to frequency weighting as the fact that the optimal rule produces lower overall variance than the rule of thumb does not mean that this superiority is uniform across frequencies. For the case where the benchmark model is the true one, i.e. $\theta = .5$, the superior performance of the optimal rule when $\frac{b_B}{b_F} = .1$ is due to the variance reductions it produces in high frequencies whereas for $\frac{b_B}{b_F} = 1$ the superior performance is due to the variance reductions in the lower frequencies. In both cases, there are intervals of frequencies where the rule of thumb outperforms the optimal rule. This illustrates how model uncertainty and preference uncertainty with respect to frequency weightings make it difficult to compare rules. The Figures also compare rule performance for $\theta = .25$ and $\theta = .75$. Overall, the relative performance of the optimal rule and rule of thumb when compared across frequencies are similar to what was observed for $\theta = .5$; this is so even though the spectral densities for $\frac{b_B}{b_F} = .1$ are shaped quite differently from the case for the benchmark model.

From the perspective of this exercise, we conclude that the case for the benchmark optimal with respect over a rule of thumb is moderately strong in the presence of model uncertainty concerning the role of forward versus backwards looking elements, if one takes a strong stand that the loss function is the overall variance of the state. However, the case is far weaker when allows for the possibility that the policymaker put different weights on different frequency-specific fluctuations.

5. Taylor versus optimal rules in an IS/PC framework

In this section we apply our various ideas to the evaluation of the Taylor rule versus an optimal rule for a two equation IS/PC system. TO BE COMPLETED

6. Summary and conclusions

Our basic conclusion is that the support for optimal rules versus Taylor's original rule is ambiguous when one considers the performance of the rules in the presence of model uncertainty and with respect to frequency-specific effects. Further we observe that the potential for optimal rules to go awry seems greater, especially for particular frequency-specific intervals. This suggests to us that great caution should be taken in seriously advocating inflation targeting, at least on the basis that they are optimal with respect to a particular dynamic programming problem. While we are not so nihilistic as to believe such knowledge is impossible, we believe that it is often the case that contemporaneous scholarly discussions of macroeconomic policy pay too little attention to Friedman's 1948 arguments. Indeed, our analysis is more interventionist than his in that our simple rules are versions of leaning against the wind policies, which Friedman specifically questions because of the problem of long and variable lags in policy effects. We do not think it unfair to say that modern time series analysis has led to a more optimistic view of the information available to policymakers than assumed by Friedman.[16] Hence, we are comfortable with recommendations that are more interventionist than his.[17]

A major weakness in our analysis is that we focus on comparisons of permanent rule choices. In other words, we ask whether a policymaker should permanently choose one rule or another. While this is the standard procedure in most of the monetary rules literature, the analysis fails to address how a rule should be chosen when the policymaker has the option of changing it in response to new information that affects the model space under consideration. One solution is suggested by Svensson and Williams (2005) who

[16] In fact, Friedman (1972) acknowledges some progress of this type, although worry about politicization preserves his skepticism of countercyclical monetary policy.

[17] As such, we are sympathetic to the common sense spirit of Blinder (1998) though in our parlance a rule can react to the state of the economy.

treat different models as regimes across which the economy switches; in their analysis the model uncertainty facing a policymaker is equivalent to uncertainty about the regime in which the economy is currently in as the conditional probabilities of future regimes given the current regime state. This is a promising research direction and admits progress using Markov jump process methods, but it treats model uncertainty in a very different way than we conceptualize it. In our approach, model uncertainty represents something that may be resolved over time, not something that exists in a steady state. Hence the next step in the approach we take seems likely to draw from ideas from the theory of bandit problems rather than Markov jump processes. And of course both approaches to model uncertainty cannot address the issue of policy evaluation when new elements emerge in the model space over time. This problem begins to link monetary policy evaluation with issues that lie at the frontiers of the work on decisionmaking under ambiguity, where even the most advanced treatment typically assumes that while the probabilities are not available for the object of interest, its support is known. So, we see this paper as only a first step in a long research program.

Table 1

Panel A: Losses Over the Model Space $a \in \left(\underline{a} = 0.05, \bar{a} = 0.95\right)$

		Total Variance	Low Freq. $\left[0, \frac{\pi}{4}\right]$	Middle-Low $\left[\frac{\pi}{4}, \frac{\pi}{2}\right]$	Middle-High $\left[\frac{\pi}{2}, \frac{3\pi}{4}\right]$	High $\left[\frac{3\pi}{4}, \pi\right]$
$\frac{\underline{a}+\bar{a}}{2}$	Min Loss	1.00	0.12	0.17	0.17	0.12
	Max Loss	**1.25**	**0.66**	**0.30**	**0.30**	**0.66**
	Average Loss	1.08	0.29	0.24	0.24	0.29
$a_1 = .3$	Min Loss	1.00	0.17	0.20	0.13	0.10
	Max Loss	**1.73**	**1.21**	**0.30**	**0.29**	**0.41**
	Average Loss	1.15	0.46	0.27	0.21	0.20
$a_1 = .8$	Min Loss	1.00	0.09	0.12	0.22	0.19
	Max Loss	**2.29**	**0.33**	**0.28**	**0.30**	**1.80**
	Average Loss	1.25	0.17	0.19	0.28	0.61

Panel B: Minimax (M) and Minimax Regret (MR)

$Weight\left(\frac{\underline{a}+\bar{a}}{2}\right) = 1-\delta$	Total Variance		Low Freq. $\left[0, \frac{\pi}{4}\right]$		Middle-Low $\left[\frac{\pi}{4}, \frac{\pi}{2}\right]$		Middle-High $\left[\frac{\pi}{2}, \frac{3\pi}{4}\right]$		High $\left[\frac{3\pi}{4}, \pi\right]$	
	M	MR	M	MR	M	MR	M	MR	M	MR
$\delta\left(a_1 = .3\right)$	**0**	**.26**	**0**	**.31**	**0**	**.42**	**1**	**.58**	**1**	**.66**
$\delta\left(a_1 = .8\right)$	**0**	**.16**	**1**	**.73**	**1**	**.62**	**0**	**.37**	**0**	**.22**

Table 2: $b_B / b_F = 0.1$

Panel A: Loss for Backward Model Under Four Alternative Policies,

	Total Variance	Low Freq. $[0, \pi/4]$	Middle-Low $[\pi/4, \pi/2]$	Middle-High $[\pi/2, 3\pi/4]$	High $[3\pi/4, \pi]$
O_B	1.00	0.25	0.25	0.25	0.25
O_F	3.06	2.57	0.31	0.11	0.08
RO_F	3.23	2.78	0.26	0.11	0.08
RoT	2.50	2.03	0.27	0.12	0.08

Panel B: Loss for Forward Model Under Four Alternative Policies

	Total Variance	Low Freq. $[0, \pi/4]$	Middle-Low $[\pi/4, \pi/2]$	Middle-High $[\pi/2, 3\pi/4]$	High $[3\pi/4, \pi]$
O_B	∞	n.a.	n.a.	n.a.	n.a
O_F	0.17	0.01	0.02	0.06	0.08
RO_F	0.63	0.06	0.08	0.15	0.35
RoT	0.81	0.03	0.04	0.10	0.64

Panel C: Minimax (M) and Minimax Regret (MR)

$Weight(O_F) = 1 - \delta$	Total Variance		Low Freq. $[0, \pi/4]$		Middle-Low $[\pi/4, \pi/2]$		Middle-High $[\pi/2, 3\pi/4]$		High $[3\pi/4, \pi]$	
	M	MR	M	MR	M	MR	M	MR	M	MR
$\delta(O_B)$	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
$\delta(RO_F)$	**0**	**.44**	**0**	**.47**	**1**	**.49**	**0**	**0**	**0**	**0**
$\delta(RoT)$	**1**	**.49**	**1**	**.56**	**1**	**.52**	**0**	**.18**	**0**	**0**

Table 3: $b_B / b_F = 1$

Panel A: Loss for Backward Model Under Four Alternative Policies,

	Total Variance	Low Freq. $[0, \pi/4]$	Middle-Low $[\pi/4, \pi/2]$	Middle-High $[\pi/2, 3\pi/4]$	High $[3\pi/4, \pi]$
O_B	1.00	0.25	0.25	0.25	0.25
O_F	3.71	0.15	3.15	0.28	0.14
RO_F	1.05	0.38	0.28	0.21	0.18
RoT	1.14	0.14	0.18	0.30	0.52

Panel B: Loss for Forward Model Under Four Alternative Policies

	Total Variance	Low Freq. $[0, \pi/4]$	Middle-Low $[\pi/4, \pi/2]$	Middle-High $[\pi/2, 3\pi/4]$	High $[3\pi/4, \pi]$
O_B	0.65	0.04	0.06	0.13	0.42
O_F	0.17	0.00	0.03	0.06	0.08
RO_F	0.63	0.06	0.08	0.15	0.35
RoT	0.81	0.03	0.04	0.10	0.64

Panel C: Minimax (M) and Minimax Regret (MR)

$\Pr(O_F) = 1 - \delta$	Total Variance		Low Freq. $[0, \pi/4]$		Middle-Low $[\pi/4, \pi/2]$		Middle-High $[\pi/2, 3\pi/4]$		High $[3\pi/4, \pi]$	
	M	MR	M	MR	M	MR	M	MR	M	MR
$\delta(O_B)$	**1**	**.79**	**0**	**.29**	**1**	**.93**	**1**	**.44**	**0**	**0**
$\delta(RO_F)$	**1**	**.78**	**0**	**.19**	**1**	**.92**	**1**	**.46**	**0**	**0**
$\delta(RoT)$	**1**	**.75**	**1**	**.46**	**1**	**.95**	**0**	**.43**	**0**	**0**

Table 4

Panel A: Losses Over the Model Space $\theta \in (0,1)$**,** $b_B / b_F = .1$

		Total Variance	Low Freq. $[0, \pi/4]$	Middle-Low $[\pi/4, \pi/2]$	Middle-High $[\pi/2, 3\pi/4]$	High $[3\pi/4, \pi]$
OB $(\theta = .5)$	Min Loss	0.78	0.01	0.03	0.10	0.09
	Max Loss	**1.88**	**1.35**	**0.42**	**0.29**	**0.79**
	Average Loss	0.99	0.22	0.20	0.20	0.37
RoT $(\rho = -1.25)$	Min Loss	0.76	0.03	0.04	0.10	0.09
	Max Loss	**2.40**	**1.91**	**0.35**	**0.25**	**0.63**
	Average Loss	1.07	0.37	0.19	0.19	0.33

Panel B: Minimax (M) and Minimax Regret (MR)

$\Pr(OB) = 1 - \delta$	Total Variance		Low Freq. $[0, \pi/4]$		Middle-Low $[\pi/4, \pi/2]$		Middle-High $[\pi/2, 3\pi/4]$		High $[3\pi/4, \pi]$	
	M	MR	M	MR	M	MR	M	MR	M	MR
$\delta(RoT)$	**0**	**.41**	**0**	**.41**	**1**	**.54**	**1**	**.55**	**1**	**.57**

Table 5

Panel A: Losses Over the Model Space $\theta \in (0,1), b_B / b_F = 1$

		Total Variance	Low Freq. $[0, \pi/4]$	Middle-Low $[\pi/4, \pi/2]$	Middle-High $[\pi/2, 3\pi/4]$	High $[3\pi/4, \pi]$
OB $(\theta = .5)$	Min Loss	0.61	0.03	0.06	0.15	0.21
	Max Loss	**1.05**	**0.18**	**0.37**	**0.32**	**0.37**
	Average Loss	0.84	0.08	0.18	0.25	0.33
RoT $(\rho = -1.25)$	Min Loss	0.81	0.03	0.04	0.10	0.52
	Max Loss	**1.14**	**0.14**	**0.18**	**0.30**	**0.68**
	Average Loss	0.99	0.07	0.09	0.19	0.64

Panel B: Minimax (M) and Minimax Regret (MR)

$\Pr(OB) = 1 - \delta$	Total Variance		Low Freq. $[0, \pi/4]$		Middle-Low $[\pi/4, \pi/2]$		Middle-High $[\pi/2, 3\pi/4]$		High $[3\pi/4, \pi]$	
	M	MR	M	MR	M	MR	M	MR	M	MR
$\delta(RoT)$	**0**	**.31**	**1**	**.58**	**1**	**.73**	**1**	**.59**	**0**	**0**

Figure 1

Backward and Forward Spectral Densities Under Alternative Policies $b_B / b_F = .1$

(Notice that under the optimal backward rule the forward model is non-stationary so the spectrum is not reported)

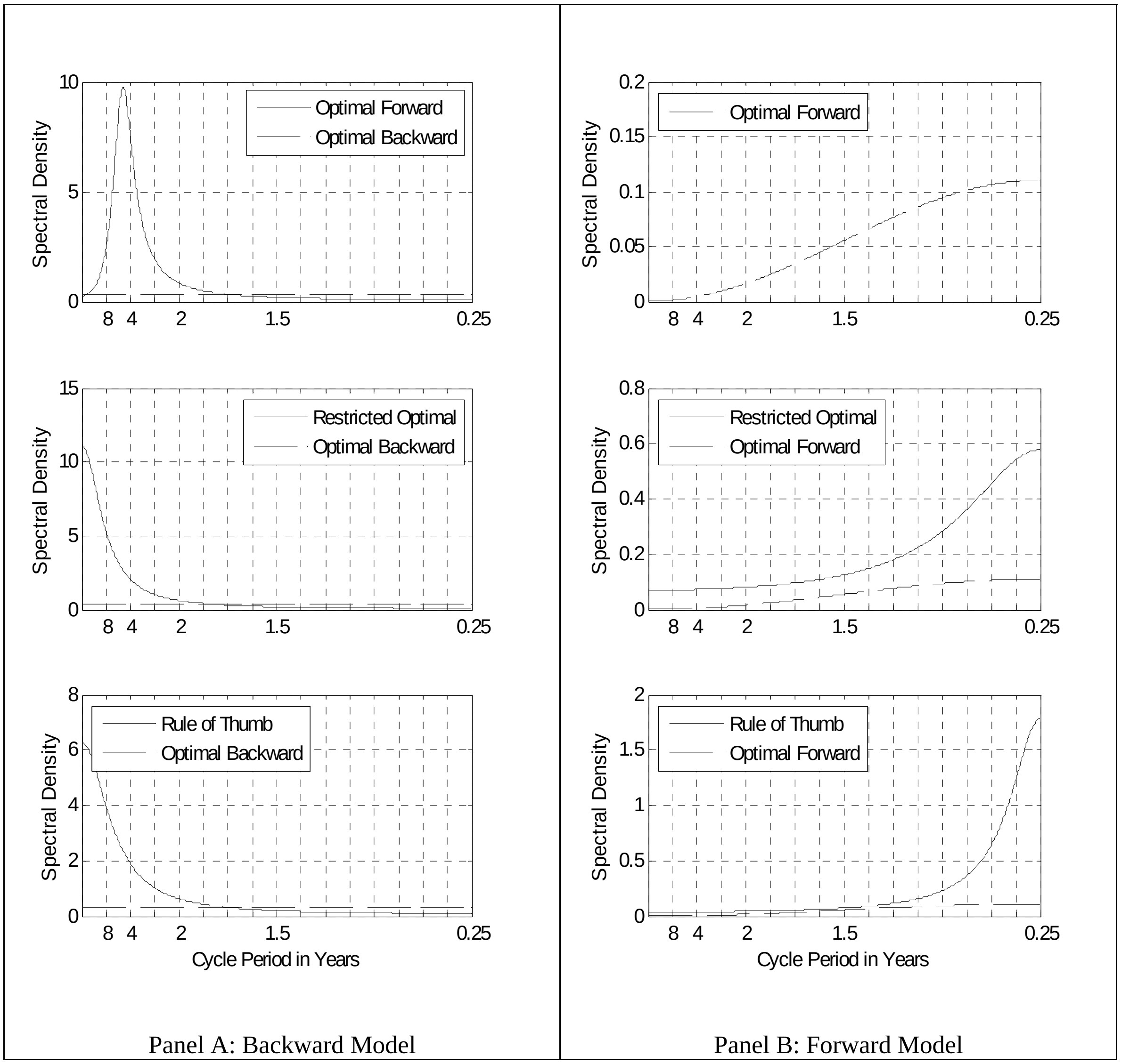
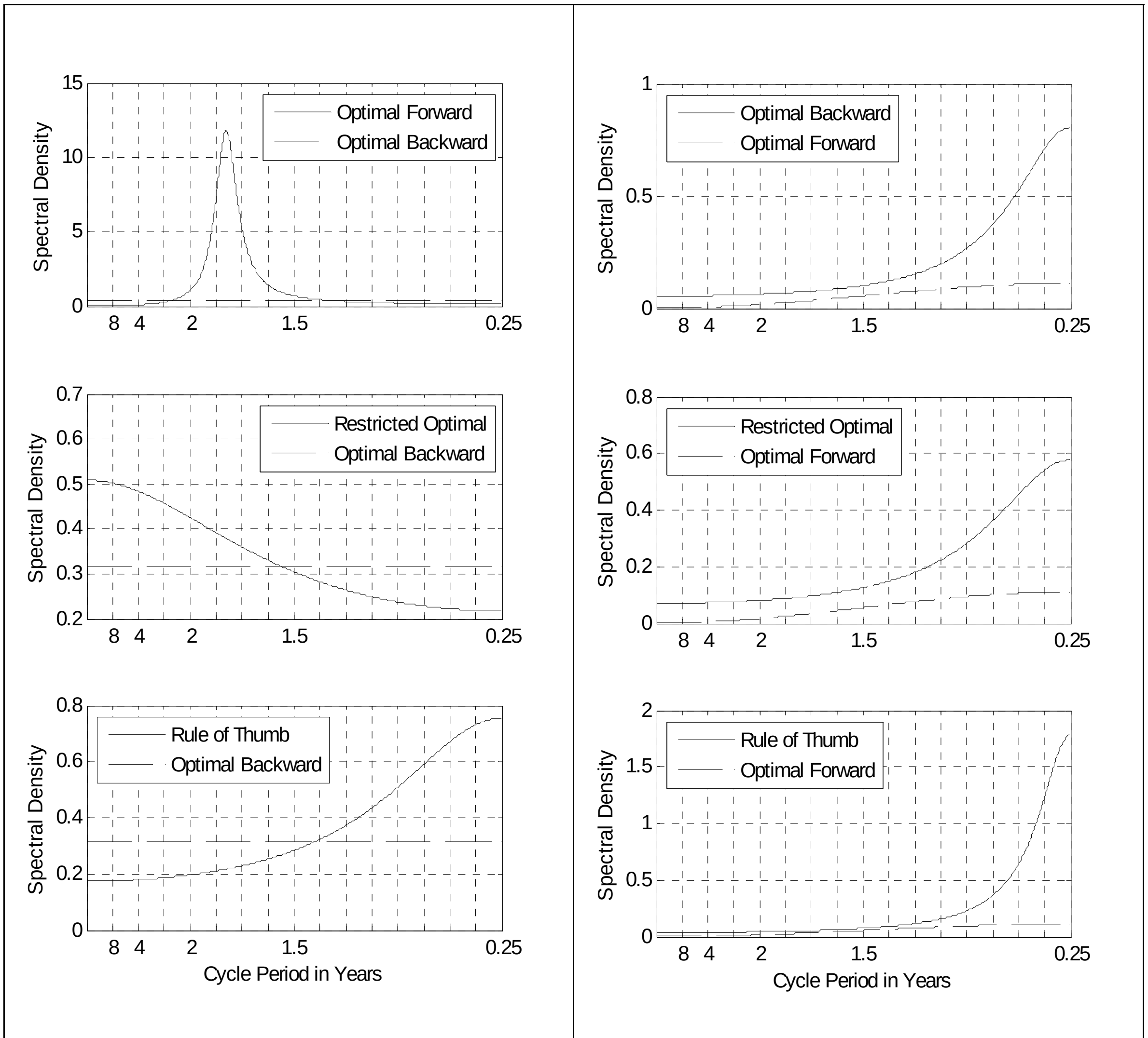


Figure 2

Backward and Forward Spectral Densities Under Alternative Policies $b_B / b_F = 1$



Panel A: Backward Model

Panel B: Forward Model

Figure 3. Frequency Specific Losses and Bode Constraint

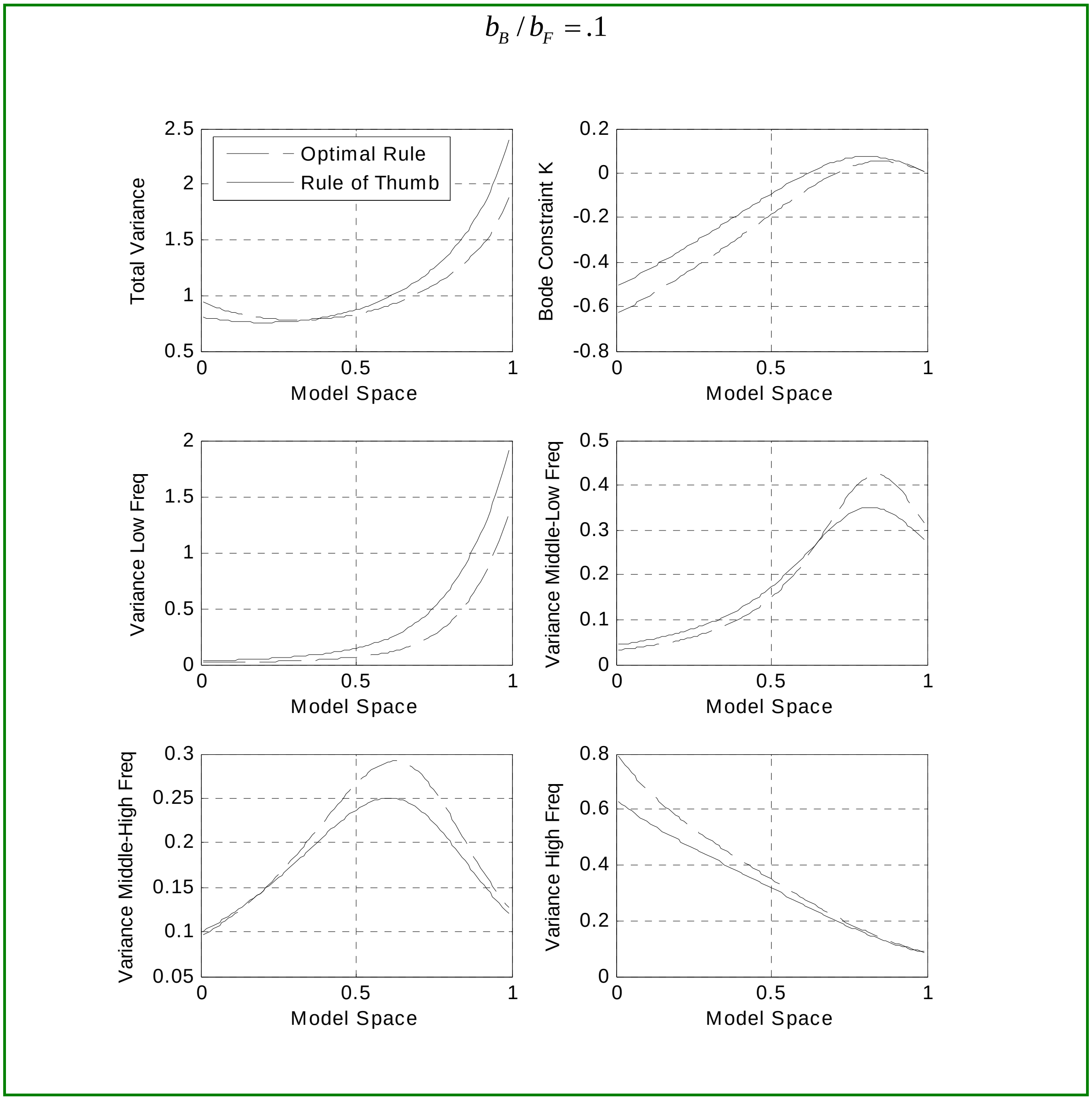


Figure 4 Frequency Specific Losses and Bode Constraint

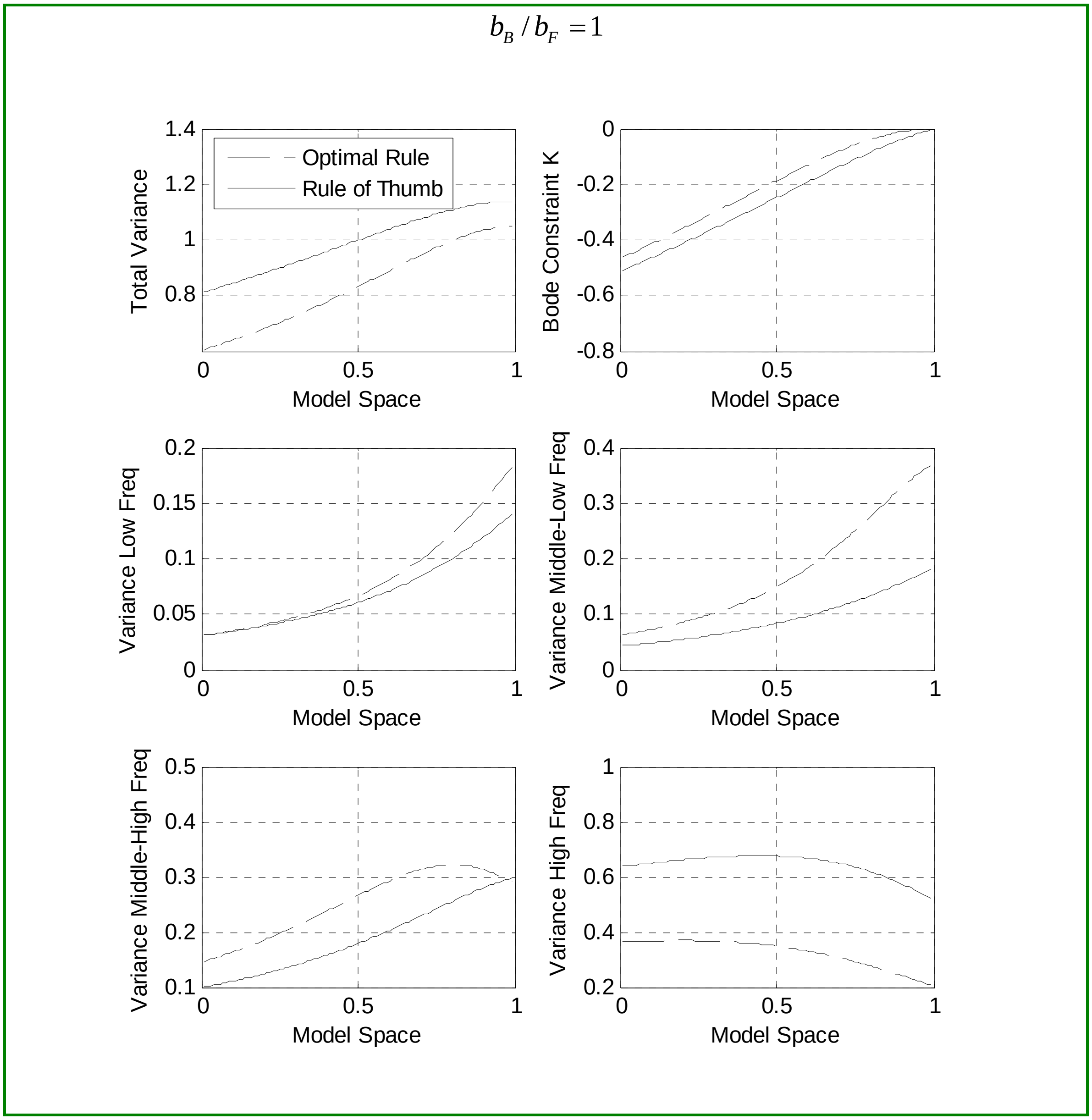
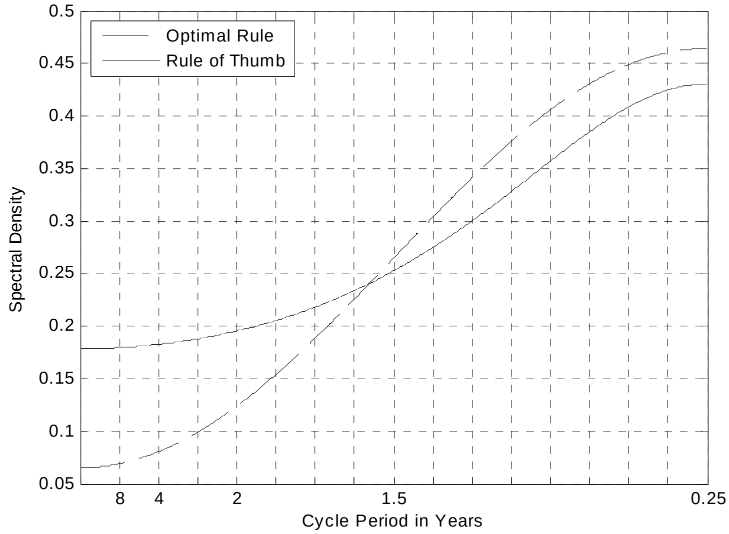
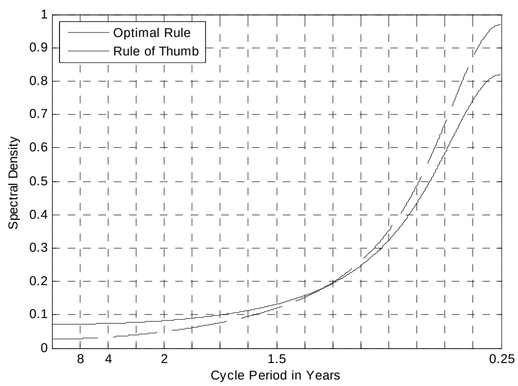


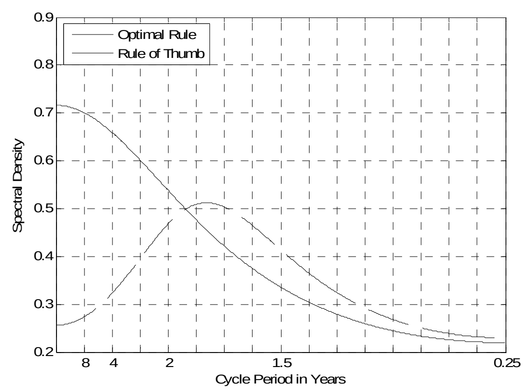
Figure 5: Variance Decomposition Under Optimal Rule for Benchmark Model
($b_B / b_F = 0.1$)



Panel A: Benchmark Model ($\theta = .5$)

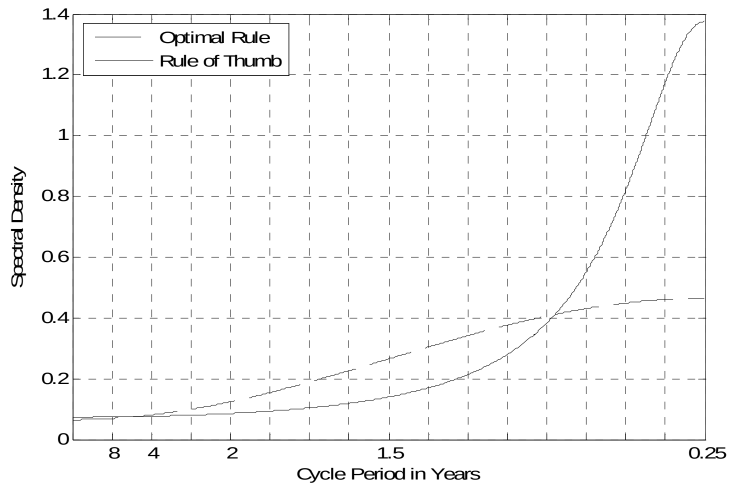


Panel B.1: $\theta = .25$

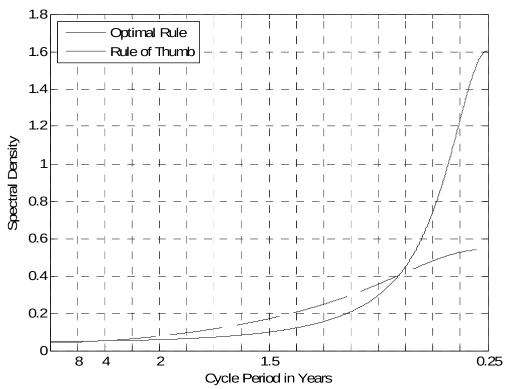


Panel B.2: $\theta = .75$

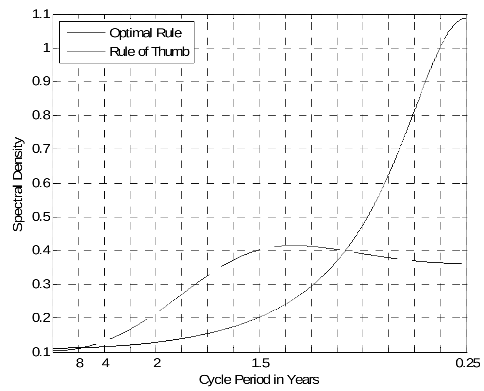
Figure 6: Variance Decomposition Under Optimal Rule for Benchmark Model ($b_B / b_F = 1.0$)



Panel A: Benchmark Model ($\theta = .5$)



Panel B.1: $\theta = .25$



Panel B.2: $\theta = .75$

Technical Appendix

Section 4.a: Derivation of minimax regret policies for optimal and rule of thumb policies under parameter uncertainty

In calculating the minimax regret solution when the candidate policies are a_1 and $a_2 = \frac{\underline{a}+\overline{a}}{2}$, the maximum regret $MR\left(a_1\right)$ will depend on the value of a_1 with respect to the alternative policy. One therefore needs to consider two cases

$$\text{Case L: } a_1 < \frac{\underline{a}+\overline{a}}{2} \Rightarrow MR\left(a_1\right) = \frac{\left(\overline{a}-a_1\right)^2}{1-\left(\overline{a}-a_1\right)^2} \tag{62}$$

and

$$\text{Case H: } \frac{\underline{a}+\overline{a}}{2} < a_1 \Rightarrow MR\left(a_1\right) = \frac{\left(\underline{a}-a_1\right)^2}{1-\left(\underline{a}-a_1\right)^2}. \tag{63}$$

In either case,

$$MR\left(a_2\right) = \left(\frac{1}{1-\left(\frac{\underline{a}-\overline{a}}{2}\right)^2}\right) - 1 = \frac{\left(\underline{a}-\overline{a}\right)^2}{4-\left(\underline{a}-\overline{a}\right)^2}. \tag{64}$$

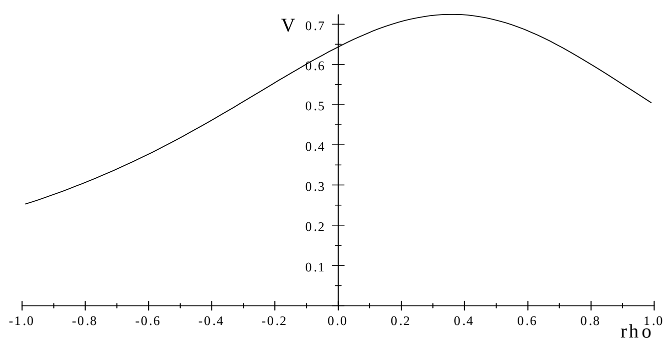
It follows that the minimax regret rules are different across the two cases, which are given in eq. (27).

Section 4.a: Frequency-specific losses

When a feedback rule induces an AR coefficient of ρ, the associated loss function is

$$\int_{\omega_L}^{\omega_H} \frac{1}{1+\rho^2 - 2\rho\cos\omega} d\omega = \frac{2}{1-\rho^2}\left(\arctan\left(\frac{1+\rho}{1-\rho}\tan\frac{\omega_H}{2}\right) - \arctan\left(\frac{1+\rho}{1-\rho}\tan\frac{\omega_L}{2}\right)\right). \quad (65)$$

The frequency interval considered in the text ($\omega_H = \frac{\pi}{2}, \omega_L = \hat{\omega}$) and assumption that that $\tan\left(\frac{\pi}{4}\right) = 1$ gives the expression in the text. The generic shape of the function is:



Section 4.b: Derivation of optimal rule in backwards looking model

The optimal policy for (38) is determined via the Lagrangean

$$\Im = E\sum_{t=0}^{\infty} \delta^t \left[\frac{1}{2}x_t^2 + \Xi_t\left(x_t - a x_{t-1} - b u_{t-1}\right)\right]. \quad (66)$$

The first order conditions are

$$\begin{cases} x_t + \Xi_t - \delta a E_t \Xi_{t+1} = 0 \\ -\delta b_B E_t \Xi_{t+1} = 0 \end{cases} \quad (67)$$

It follows that

$$x_t = -\Xi_t \tag{68}$$

and so

$$E_t x_{t+1} = -E_t \Xi_{t+1} = 0 \,. \tag{69}$$

Applying the expectations operator it is evident that

$$\begin{aligned} E_t x_{t+1} &= a x_t + b_B u_t + E_t \varepsilon_{t+1} \\ 0 &= a x_t + b_B u_t \end{aligned} \tag{70}$$

which leads to the optimal rule backward looking rule reported in the text.

Section 4.b: Derivation of Optimal Forward Looking Rule

The Lagrangean for (38) is

$$\Im = E\sum_{t=0}^{\infty} \delta^t \left[\frac{1}{2} x_t^2 + \Xi_t \left(x_t - \beta x_{t+1} - b u_{t-1} \right) \right] \tag{71}$$

which produces first order conditions

$$\begin{cases} x_t + \Xi_t - \dfrac{\beta}{\delta} \Xi_{t-1} = 0 \\ -\delta b_F E_t \Xi_{t+1} = 0 \end{cases} \tag{72}$$

so that

$$x_t = -\Xi_t\left(1-\frac{\beta}{\delta}L\right). \tag{73}$$

Leading the forward model one period and substituting

$$E_t x_{t+1} = -E_t\Xi_{t+1}\left(1-\frac{\beta}{\delta}L\right) = -E_t\Xi_{t+1} + \frac{\beta}{\delta}\Xi_t = \frac{\beta}{\delta}\Xi_t . \tag{74}$$

Substituting into the Lagrangean constraint

$$-\Xi_t\left(1-\frac{\beta}{\delta}L\right) = \frac{\beta^2}{\delta}\Xi_t + b_F u_{t-1} + \varepsilon_t . \tag{75}$$

Leading this forward one period and taking expectations

$$\frac{\beta}{\delta}\Xi_t = b_F u_t \Rightarrow \ \Xi_t = \frac{\delta}{\beta} b_F u_t \tag{76}$$

so that

$$x_t = -\frac{\delta}{\beta} b_F \left(1-\frac{\beta}{\delta}L\right) u_t \tag{77}$$

which leads to the expression in the text.

Section 4.c: Fully optimal rule for a hybrid model

The (unconditional) Lagrangean for this problem is

$$\Im = E\sum_{t=0}^{\infty}\delta^t\left[\frac{1}{2}x_t^2 + \Xi_t\left(x_t - (1-\theta)\beta x_{t+1} - \theta a x_{t-1} - b(\theta)u_{t-1}\right)\right] \tag{78}$$

where

$$b(\theta) = \theta b_B + (1-\theta) b_F . \tag{79}$$

The set of first order conditions are

$$\begin{cases} x_t + \Xi_t - (1-\theta)\dfrac{\beta}{\delta}\Xi_{t-1} - \theta a \delta E_t \Xi_{t+1} = 0 \\ -\delta b(\theta) E_t \Xi_{t+1} = 0 \end{cases} \tag{80}$$

It immediately follows that

$$x_t = -\Xi_t \left(1 - (1-\theta)\frac{\beta}{\delta} L \right) \tag{81}$$

and so

$$E_t x_{t+1} = -E_t \left[\Xi_{t+1} \left(1 - (1-\theta)\frac{\beta}{\delta} L \right) \right] = (1-\theta)\frac{\beta}{\delta}\Xi_t . \tag{82}$$

Substituting the expressions for x_t and into the hybrid constraint one finds the relationship that the multiplier and the instrument must obey on an optimal path,

$$-\Xi_t \left(1 - (1-\theta)\frac{\beta}{\delta} L \right) = -(1-\theta)\frac{\beta^2}{\delta}\Xi_t - \theta a \Xi_{t-1} \left(1 - (1-\theta)\frac{\beta}{\delta} L \right) + b(\theta) u_{t-1} + \varepsilon_t . \tag{83}$$

Leading this equation forward one period and applying the expectation operator to it produces

$$\left(1-\theta\right)\frac{\beta}{\delta}\Xi_t = -\theta a \Xi_t\left(1-\left(1-\theta\right)\frac{\beta}{\delta}L\right)+b\left(\theta\right)u_t \tag{84}$$

rearranging the instrument and the multiplier are related by:

$$\begin{aligned} u_t &= \\ &\frac{1}{b\left(\theta\right)}\left(\left(1-\theta\right)\frac{\beta}{\delta}+\theta a-\theta\left(1-\theta\right)\frac{a\beta}{\delta}L\right)\Xi_t = \\ &\frac{1}{b\left(\theta\right)}\left(\frac{\left(1-\theta\right)\beta+\delta\theta a}{\delta}-\theta\left(1-\theta\right)\frac{a\beta}{\delta}L\right)\Xi_t = \\ &\frac{\left(1-\theta\right)\beta+\delta\theta a}{b\left(\theta\right)\delta}\left(1-\frac{\theta\left(1-\theta\right)a\beta}{\left(1-\theta\right)\beta+\delta\theta a}L\right)\Xi_t. \end{aligned} \tag{85}$$

Notice that as $\theta \to 0$ this condition delivers the relationship between the multiplier and the instrument for the pure forward looking case.

We next consider the optimal relationship between the instrument and the state which is bridged by the process for the multiplier. Let $\alpha\left(\theta\right) \equiv -\frac{\left(1-\theta\right)\beta+\delta\theta a}{\delta}$, $\gamma\left(\theta\right) \equiv \left(1-\theta\right)\frac{\beta}{\delta}$, $B\left(\theta\right) \equiv \frac{\theta\left(1-\theta\right)a\beta}{\left(1-\theta\right)\beta+\delta\theta a}$ and assume that $\left|\xi\left(\theta\right)\right|<1$ so that we can write the multiplier in terms of the instrument and substituting

$$x_t = \frac{b\left(\theta\right)\left(1-\gamma\left(\theta\right)L\right)}{\alpha\left(\theta\right)\left(1-B\left(\theta\right)L\right)}u_t \tag{86}$$

from which, under the assumption $\left|\left(1-\theta\right)\frac{\beta}{\delta}\right|<1$, one can derive the optimal instrument rule

$$u_t^*(\theta)=\frac{\alpha(\theta)}{b(\theta)}\frac{\left(1-B(\theta)L\right)}{\left(1-\gamma(\theta)L\right)}x_t =$$

$$-\frac{(1-\theta)\beta+\delta\theta a}{b(\theta)\delta}\frac{\left(1-\frac{\theta(1-\theta)a\beta}{(1-\theta)\beta+\delta\theta a}L\right)}{\left(1-(1-\theta)\frac{\beta}{\delta}L\right)}x_t. \tag{87}$$

An instrument rule combined with the law of motion for the state allows to write an expectational difference equation only in terms of the state. In what follows we solve for the path of the state variable x_t by applying the instrument rule for a benchmark model $\hat{\theta}$, $u_t^*\left(\hat{\theta}\right)=\frac{\alpha\left(\hat{\theta}\right)}{b\left(\hat{\theta}\right)}\frac{\left(1-B\left(\hat{\theta}\right)L\right)}{\left(1-\gamma\left(\hat{\theta}\right)L\right)}x_t$, to a law of motion of a generic model θ which, namely we want to solve for

$$x_t=(1-\theta)\beta E_t x_{t+1}+\theta a x_{t-1}+b(\theta)\frac{\alpha\left(\hat{\theta}\right)}{b\left(\hat{\theta}\right)}\frac{\left(1-B\left(\hat{\theta}\right)L\right)}{\left(1-\gamma\left(\hat{\theta}\right)L\right)}x_{t-1}+\varepsilon_t. \tag{88}$$

Multiply both sides of (88) by $\left(1-\gamma\left(\hat{\theta}\right)L\right)$ so that the equation becomes

$$\begin{aligned}\left(1-\gamma\left(\hat{\theta}\right)L\right)x_t&=\left(1-\gamma\left(\hat{\theta}\right)L\right)(1-\theta)\beta E_t x_{t+1}+\left(1-\gamma\left(\hat{\theta}\right)L\right)\theta a x_{t-1}\\&+b(\theta)\frac{\alpha\left(\hat{\theta}\right)}{b\left(\hat{\theta}\right)}\left(1-B\left(\hat{\theta}\right)L\right)x_{t-1}+\left(1-\gamma\left(\hat{\theta}\right)L\right)\varepsilon_t\end{aligned} \tag{89}$$

or

$$x_t = (1-\theta)\beta E_t x_{t+1} - \gamma(\hat{\theta})(1-\theta)\beta E_{t-1} x_t + \left(\gamma(\hat{\theta}) + \theta a + b(\theta)\frac{\alpha(\hat{\theta})}{b(\hat{\theta})}\right) x_{t-1}$$
$$-\left(\gamma(\hat{\theta})\theta a + b(\theta)\frac{\alpha(\hat{\theta})}{b(\hat{\theta})} B(\hat{\theta})\right) x_{t-2} + \left(1-\gamma(\hat{\theta})L\right)\varepsilon_t. \tag{90}$$

Let $H(\theta,\hat{\theta}) = \left(\gamma(\hat{\theta}) + \theta a + b(\theta)\frac{\alpha(\hat{\theta})}{b(\hat{\theta})}\right)$ and $K(\theta,\hat{\theta}) = -\gamma(\hat{\theta})\theta a - b(\theta)\frac{\alpha(\hat{\theta})}{b(\hat{\theta})} B(\hat{\theta})$.

It can be shown that when the optimal rule is applied to the appropriate model these two objects take equal zero, i.e.

$$H(\theta,\theta) = \gamma(\theta) + \theta a + b(\theta)\frac{\alpha(\theta)}{b(\theta)} =$$
$$(1-\theta)\frac{\beta}{\delta} + \theta a - \frac{(1-\theta)\beta + \delta\theta a}{\delta} = 0 \tag{91}$$

and

$$K(\theta,\theta) = -\gamma(\theta)\theta a - b(\theta)\frac{\alpha(\theta)}{b(\theta)} B(\theta) =$$
$$-(1-\theta)\frac{\beta}{\delta}\theta a + \frac{(1-\theta)\beta + \delta\theta a}{\delta}\frac{\theta(1-\theta)a\beta}{(1-\theta)\beta + \delta\theta a} = 0. \tag{92}$$

This shows that when the optimal policy is tailored to the wrong model higher order dynamics are activated in the law of motion for the state. We solve the expectational difference equation by guessing a square summable moving average form

$$x_t = G_{\theta,\hat{\theta}}(L)\varepsilon_t. \tag{93}$$

Dropping the subscripts for notational convenience we have

$$\begin{aligned} G(L)\varepsilon_t &= \\ (1-\theta)\beta\left(G(L)-G_0\right)L^{-1}\varepsilon_t &- \gamma\left(\hat{\theta}\right)(1-\theta)\beta\left(G(L)-G_0\right)\varepsilon_t \\ +H\left(\theta,\hat{\theta}\right)G(L)L\varepsilon_t &+ K\left(\theta,\hat{\theta}\right)G(L)L^2\varepsilon_t + \left(1-\gamma\left(\hat{\theta}\right)L\right)\varepsilon_t. \end{aligned} \tag{94}$$

Multiplying both sides by L, rearranging and making use of the Riesz-Fischer theorem one can write

$$G(L) = \frac{\left(1-\gamma\left(\hat{\theta}\right)L\right)\left(G_0(1-\theta)\beta - L\right)}{\left((1-\theta)\beta - \left(1+(1-\theta)\beta\gamma\left(\hat{\theta}\right)\right)L + H\left(\theta,\hat{\theta}\right)L^2 + K\left(\theta,\hat{\theta}\right)L^3\right)}. \tag{95}$$

To ensure a unique solution exactly one root of the denominator must be inside the unit circle (i.e. the dominant eigenvalue bigger than one) so that G_0 can be chosen in order to exactly cancel this unstable root. The denominator can be always represented as:

$$(1-\theta)\beta\left(1-\lambda_1\left(\theta,\hat{\theta}\right)L\right)\left(1-\lambda_2\left(\theta,\hat{\theta}\right)L\right)\left(1-\lambda_3\left(\theta,\hat{\theta}\right)L\right). \tag{96}$$

In what follows we assume that for any combination we consider we have

$$\left|\lambda_1\left(\theta,\hat{\theta}\right)\right| > 1, \left|\lambda_2\left(\theta,\hat{\theta}\right)\right|, \left|\lambda_3\left(\theta,\hat{\theta}\right)\right| < 1. \tag{97}$$

If this is not satisfied there exists a set of solutions that satisfy the first order conditions. The unique solution is identified by the condition

$$\left(G_0\left(1-\theta\right)\beta-\frac{1}{\lambda_1\left(\theta,\hat{\theta}\right)}\right)=0\Rightarrow G_0=\frac{1}{\lambda_1\left(\theta,\hat{\theta}\right)\left(1-\theta\right)\beta}. \tag{98}$$

Finally, substituting in the generic solution and canceling the roots the solution becomes

$$G_{\theta,\hat{\theta}}\left(L\right)=\frac{\left(1-\gamma\left(\hat{\theta}\right)L\right)}{\lambda_1\left(\theta,\hat{\theta}\right)\left(1-\theta\right)\beta\left(1-\lambda_2\left(\theta,\hat{\theta}\right)L\right)\left(1-\lambda_3\left(\theta,\hat{\theta}\right)L\right)}. \tag{99}$$

When the optimal rule is applied to the appropriate model the higher order dynamics disappear and the solution becomes a simple MA(1)

$$G_{\hat{\theta},\hat{\theta}}\left(L\right)=\frac{\left(1-\gamma\left(\hat{\theta}\right)L\right)}{\lambda_1\left(\hat{\theta},\hat{\theta}\right)\left(1-\hat{\theta}\right)\beta}. \tag{100}$$

The normalized spectral densities reported in the main text are then computed by evaluating the moving average on the unit circle, i.e. $L=e^{-i\omega}$,

$$f_x\left(\omega\right)=\frac{1}{2\pi}\sigma_\varepsilon^2 G\left(e^{-i\omega}\right)G\left(e^{i\omega}\right). \tag{101}$$

Bibliography

Blinder, A. 1998. *Central Banking in Theory and Practice*. Cambridge: MIT Press.

Blume, L., D. Easley, and J. Halperin. 2006. Redoing the Foundations of Decision Theory. Mimeo, Cornell University.

Brock. W. 2006. Profiling Problems With Partially Identified Structure. *Economic Journal*, forthcoming.

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

Brock, W. and S. Durlauf, (2005), "Local Robustness Analysis: Theory and Application," *Journal of Economic Dynamics and Control* 29, 11, 2067-2092.

Brock, W., S. Durlauf, and G. Rondina. 2006. Design Limits and Dynamic Policy Analysis. Mimeo, University of Wisconsin.

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

Brock, W., S. Durlauf, K. West. 2006. Model Uncertainty and Policy Evaluation: Some Theory and Empirics. *Journal of Econometrics,* forthcoming.

Camerer, C. and G. Loewenstein. 2003. Behavioral Economics: Past, Present, Future. "is in *Advances in Behavioral Economics*, C. Camerer, G. Loewenstein, and M. Rabin, eds. Princeton: Princeton University Press.

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

Chernoff, H. 1954. Rational Selection of Decision Functions. *Econometrica* 22: 422-443

Clarida, R., J. Gali and M. Gertler, 1999. The Science of Monetary Policy: A New Keynesian Perspective. *Journal of Economic Literature* XXXVII(December), 1661-1701

Epstein, L. and T. Wang. 1994. Intertemporal Asset Pricing Behavior Under Knightian Uncertainty. *Econometrica*, 62, 283-322.

Freedman, D. 1991. A Rejoinder to Berk, Blalock, and Mason. *Sociological Methodology* 21, 353-358, 1991.

Friedman, M. 1948. A Monetary and Fiscal Framework for Economic Stability. *American Economic Review* 38, 3, 245-264.

Friedman, M. 1972. Monetary Policy. *Proceedings of the American Philosophical Society* 116(3), 183-196.

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

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

Hansen, L., T. Sargent. 2003. Robust Control of Forward-Looking Models. *Journal of Monetary Economics*, 50, 581-604.

Hansen, L., T. Sargent. 2006. *Robustness.* manuscript, Hoover Institution, Stanford University and forthcoming, Princeton University Press.

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

Key, J., L. Perrichi, and A. Smith. 1998. Choosing Among Models when None of Them are True. In *Proceedings of the Workshop on Model Selection, Special Issue of Rassegna di Metodi Statistici ed Applicazioni*, W. Racugno, ed., Bologna: Pitagora Editrice.

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

Levin, A., J. Williams. 2003. Robust Monetary Policy with Competing Reference Models. *Journal of Monetary Economics,* 50, 5, 945-975.

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

Manski, C. 2006. Minimax-Regret Threatment Choice with Missing Outcome Data. *Journal of Econometrics*, forthcoming.

McCallum, B. 1988. Robustness Properties of a Rule for Monetary Policy. *Carnegie-Rochester Conference Series on Public Policy*, 29, 173-204.

McCallum, B. 1999. Issues in the Design of Monetary Policy Rules. *Handbook of Macroeconomics*, J. Taylor and B. McCallum, eds. Amsterdam: North-Holland, 1483-1530.

McCallum, B. and E. Nelson. 2005. Targeting versus Instrument Rules for Monetary Policy. *Federal Reserve Bank of St. Louis Review*, September/October, 597-611.

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.

Otrok. C. 2001. Spectral Welfare Cost Functions. *International Economic Review*, 42, 345-367.

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* 112(April), 402-432.

Rudebusch, G. and L. Svensson. 1999. Policy Rules for Inflation Targeting. In *Monetary Policy Rules*, J. Taylor, ed. Chicago: University of Chicago Press.

Rudd, J. and K. Whelan. 2005. “New Tests of the new-Keyneisan Phillips Curve,” *Journal of Monetary Economics* 52, 1167-1181.

Savage, L. J. 1951. The Theory of Statistical Decision. *Journal of the American Statistical Association*, 46, 55-67.

Stoye, J. 2006. Statistical Decisions under Ambiguity. Mimeo, New York University.
Targeting Rules. *Journal of Economic Literature*, 42, 2, 426-477.

Svensson, L. 2002. Robust Control Made Simple. Mimeo, Princeton University.

Svensson. L. 2003. What is Wrong with Taylor Rules? Using Judgment in Monetary Policy through Targeting Rules. *Journal of Economic Literature*, 42, 2, 426-477.

Svensson, L. and N. Williams. 2005. Monetary Policy with Model Uncertainty: Distribution Foreacast Targeting. Mimeo, Princeton University.

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.

Tobin, J. 1983. Monetary Policy: Rules, Targets, and Shocks. *Journal of Money, Credit, and Banking* 15(4), 506-518.

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