Modeling Model Uncertainty

Alexei Onatski
Department of Economics
Columbia University
e-mail: ao2027@columbia.edu

Noah Williams*
Department of Economics
Princeton University
e-mail: noahw@princeton.edu

June 5, 2002

Abstract

Recently there has been a great deal of interest in studying monetary policy under model uncertainty. We develop methods to analyze different sources of uncertainty in one coherent structure, which is useful for policy decisions. We show how to estimate the size of the uncertainty based on time series data, and how to incorporate this uncertainty in policy optimization. In particular, we propose two different approaches to modeling model uncertainty. The first is model error modeling, which imposes additional structure on the errors of an estimated model, and builds a statistical description of the uncertainty around a model. The second approach is set membership identification, which uses a deterministic approach to find a set of models which are consistent with the data and prior assumptions. The center of this set becomes a benchmark model, and the radius of the set is a measure of the model uncertainty. Using both approaches, we compute the robust monetary policy under different specifications of model uncertainty in a small model of the US economy.

1 Introduction

Uncertainty is pervasive in economics, and this uncertainty must be faced continually by policy makers. In this paper we propose empirical methods to specify and measure uncertainty associated with economic models, and we study the effects of uncertainty on monetary policy decisions. Recently there has been a great deal of research activity on monetary policy making under uncertainty. We add to this literature by developing new, coherent methods to quantify uncertainty and to tailor decisions to the empirically relevant sources of uncertainty. In particular, we will be concerned with four types of uncertainty: first, uncertainty about

*We thank Mario Milanese for extensive discussions, and thank Glenn Rudebusch and Athanasios Orphanides for providing us with data.

parameters of a reference model (including uncertainty about the model's order); second, uncertainty about the spectral characteristics of noise; third, uncertainty about data quality; and fourth, uncertainty about the reference model itself. We use a simple, empirically-based macroeconomic model in order to analyze the implications of uncertainty for economic policy.

A common intuitive view is that the introduction of uncertainty should make policy makers cautious. This view reflects the results of Brainard (1967), which Blinder (1997) summarizes as "Brainard's conservatism principle: estimate what you should do, and then do less." However, as was argued by Chow (1975) from the theoretical point of view, uncertainty need not result in cautiousness.[1] Recent research on the robustness of monetary policy under model and data uncertainty confirms Chow's argument quite forcefully. The basic idea is that some types of uncertainty mean that policy instruments may turn out to have weaker effects than expected, which can in turn lead to large potential losses. In these cases, it is optimal to react more aggressively with uncertainty than without.

Moreover, it seems that different assumptions about the structure of uncertainty have drastically different implications for policy activism. For example, introduction of an extreme shock uncertainty into the Ball (1999) model, so that the serial correlation structure of the shocks is not restricted in any way (see Sargent (1999)), implies an aggressive robust Taylor rule.[2] On the contrary, Rudebusch (2001) shows that focusing on the real time data uncertainty in a conceptually similar Rudebusch and Svensson (1999) model leads to attenuation of the parameters of the optimal Taylor rule. Finally, Onatski and Stock (2002) find that uncertainty about the lag structure of the Rudebusch-Svensson model requires a cautious reaction to inflation but an aggressive reaction to the output gap.

The fact that the robust policy rules are so fragile with respect to different assumptions about the structure of uncertainty is not surprising by itself. Indeed, for any given policy rule it is possible to design a set of models (specific for this particular rule) such that the rule will be the best against model uncertainty represented by this set.[3] Hence, special choices of uncertainty sets can make any given policy rule robust. Fragility is a general feature of optimizing models. As Carlson and Doyle (2002) state, "They are 'robust, yet fragile', that is, robust to what is common or anticipated but potentially fragile to what is rare or unanticipated." Standard stochastic control methods are robust to realizations of shocks, as long as the shocks come from the assumed distributions and feed through the model in the specified way. But the optimal rules may perform poorly when faced with a different shock distribution, or slight variation in the model. The Taylor policy rules discussed above are each designed to be robust to a particular type of uncertainty, but may perform poorly when faced with uncertainty of a different nature.

In our view, the most important message of the fragility of the robust rules found in the

[1]Although the conservatism result is better known, Brainard (1967) also notes that a large enough covariance between the shocks and (random) parameters can lead policy to be more aggressive.

[2]The robust rule is defined as the policy rule that minimizes losses under the worst possible scenario consistent with the uncertainty description.

[3]This argument can be made precise with the help of the Youla parameterization of all stabilizing controllers (see Zhou, Doyle, and Glover (1996)).

literature is that to design a robust policy rule in practice, it is necessary to combine different sources of uncertainty in one coherent structure and carefully estimate or calibrate the size of the uncertainty. Further, the description and analysis of uncertainty should reflect the use of the model for policy purposes. Model selection and evaluation should not be based only on the statistical fit or predictive ability of the model, but should be tied to the ultimate policy design objectives. For these reasons, it is necessary to *model* the model uncertainty.

It is sometimes argued that model uncertainty can be adequately represented by suitable restrictions on the joint distribution of shocks only. We argue that, if the model uncertainty is used to formulate the robust policy rule, the distinction between restrictions on the vector of model parameters and the distribution of shocks may be crucial. In particular, using the Rudebusch-Svensson model as our benchmark model, we develop an example showing that the Hansen and Sargent (2002) approach to formulating model uncertainty may lead to the design of robust policy rules that can be destabilized by small parametric perturbations. This potential inconsistency between the robustness to shock uncertainty and the robustness to parametric uncertainty results from the fact that the magnitude of the shock uncertainty relevant for policy evaluation cannot be judged *ex ante*, but depends on the policy being analyzed. For example, uncertainty about the slope of the IS curve is equivalent to larger shocks under more aggressive interest rate rules.

In this paper we consider two particular ways to structure the model uncertainty. Both draw upon the recent advances in control system identification literature. First, we consider the Model Error Modeling (MEM) approach as in Ljung (1999). The main idea of the approach is as follows. First, estimate a reference or nominal model. Then take the reference model's errors and try to fit them with a general set of explanatory variables (including some variables omitted from the original model). Finally, build the model uncertainty set around the reference model which is consistent with all regressions for the errors not rejected by formal statistical procedures.

The second approach taken in this paper exploits recent advances in the Set Membership (SM) identification literature, due to Milanese and Taragna (2001). The set membership literature takes a deterministic approach to model uncertainty. The uncertainty corresponds to an infinite number of linear constraints on the model's parameters and shocks. Typically, shocks are required to be bounded in absolute value by a given positive number and the model impulse responses are required to decay at a given exponential rate. The model uncertainty set is represented by those models that are not falsified by the data and the above linear restrictions. The SM approach makes very few *a priori* assumptions about the model's shocks and parameters. While this generality is a benefit of the approach, it implies that the resulting uncertainty set may be very large and complicated. To make it operational, it is modeled (approximated) by a ball in the model space that covers the uncertainty set and has minimal radius among all such balls. The center of this ball serves as the nominal model. Thus this approach provides both an estimate of a nominal model and a description of the model uncertainty, based entirely on the specified assumptions and the observed data.

After a model of the model uncertainty is built (under either approach), the robust policy is formulated so that it works well for all models described by this model uncertainty model.

In order to guarantee the *uniform* performance across the different models, following much of the recent literature, we use a minimax approach to formulating robust policies. Although an alternative Bayesian approach has strong theoretical foundations, it is less tractable computationally. Further, minimax policies have alternative theoretical foundations and they are naturally related to some of our estimation methods.[4]

Under both the model error modeling and set membership approaches, a certain level of subjectivity exists at the stage of formulating the model of model uncertainty. In the MEM approach, one has to choose a set of explanatory variables for the model of errors and the level of statistical tests rejecting those models inconsistent with the data. In the SM approach, it is necessary to specify *a priori* a bound on absolute value of the shocks and the rate of exponential decay of the impulse responses. How should one make these subjective choices? This question is not easy to answer. If one wants to decrease subjectivity, one is forced to consider more and more general models of the model errors or, in the set membership case, less and less restrictive assumptions on the rate of decay and the shock bound. However, an effect of such vagueness may be an enormous increase in the size of the uncertainty.

One possible way to judge the specification of model uncertainty models is to do some ex-post analysis and examination. The end results of our procedures are robust Taylor-type rules and guaranteed upper bounds on a quadratic loss function. One procedure for assessing the specification of model uncertainty models would be to see if the bounds on the loss function or the robust rules themselves satisfy some criterion of "reasonability." Another procedure which Sims (2001) suggests would be to examine in more detail the implied worst-case model which results from the policy optimization, to see if it implies plausible prior beliefs. By doing this ex-post analysis, the specifications could be tailored or the results from different specifications could be weighted. Our results should prepare the reader to carry out this task by incorporating his own subjective beliefs.

The general message of our results is that there is substantial uncertainty in a class of simple models used for monetary policy. Further, different ways of modeling uncertainty can lead to quite different outcomes. We first assess the different sources of uncertainty in Rudebusch and Svensson (1999) model under our Model Error Modeling approach. We find that the famous Taylor (1993) rule leads to extremely large losses for very small perturbations. Of the different sources of uncertainty, model uncertainty has the largest effect on losses, the real-time data uncertainty is less dangerous for policy making, whereas the effects of pure shock uncertainty are relatively mild. While the full estimated model of uncertainty is too large to guarantee finite losses for any Taylor-type rules, we are able to find the rules optimally robust against specific blocks of the uncertainty model taken separately. We find that almost all computed robust rules are relatively more aggressive than the optimal rule under no uncertainty. An exception to this aggressiveness result is the rule robust to the real-time data uncertainty about the output gap. This rule is substantially less aggressive than the optimal rule under no uncertainty, which accords with results of Rudebusch (2001).

Under the Set Membership approach, we assess the uncertainty of different parametric

[4]Gilboa and Schmeidler (1989) provide an axiomatic foundation for max-min expected utility. While we share a similar approach, their setting differs from ours.

models, and develop an estimation procedure that minimizes a measure of model uncertainty. We analyze two sets of *a priori* assumptions, and find that the estimates and the evaluation of uncertainty differ dramatically. In all cases, the amount of uncertainty we estimate is substantial, and tends to be concentrated at high frequencies. Further, the models we estimate imply substantially different inflation dynamics than a conventional OLS estimate. In order to obtain reasonable policy rules, we scale down the model uncertainty in the different estimates, and we find that the resulting robust Taylor rules respond more aggressively to both inflation and the output gap relative to the optimal Taylor rules under no uncertainty.[5]

Under both of our approaches the most damaging perturbations of the reference model result from very low frequency movements. Since we impose a vertical long run Phillips curve, increases in the output gap lead to very persistent increases in inflation under relatively non-aggressive interest rate rules. The size of this persistent component is poorly measured, but has a huge impact on the losses sustained by the policy maker. We believe that for practical purposes, it is prudent to downweight the importance of the low frequency movements. The baseline model that we use, due to Rudebusch and Svensson (1999), is essentially an aggregate demand-aggregate supply model. Such models are not designed to capture long-run phenomena, but are instead most appropriately viewed as short-run models of fluctuations. By asking such a simple model to accommodate very low frequency perturbations, we feel that we are pushing the model too far. A more fully developed model is necessary to capture low frequency behavior.

To tailor our uncertainty description to more relevant worst scenarios, we reconsider our results when restricting our attention to business cycle frequencies (corresponding to periods from 6 to 32 quarters). For both MEM and SM approaches, our aggressiveness result becomes reversed. Now, almost all optimally robust Taylor rules are less aggressive than the optimal Taylor rules under no uncertainty. For the MEM approach, the least aggressive robust rule corresponds to the uncertainty about the slope of the IS curve and not to the real-time data uncertainty about the output gap. The full estimated uncertainty model is still too big to allow for finite worst possible losses under any Taylor-type rule, however, when we scale the size of all uncertainty downwards, the optimally robust Taylor rule has the coefficient on inflation 1.4 and the coefficient on the output gap 0.7 which is surprisingly close to the Taylor (1993) rule! For the SM approach, our estimated model can handle the full estimated uncertainty, and it also leads to a relatively less aggressive policy rule.

Two recent studies are particularly relevant for the Model Error Modeling part of our paper. Rudebusch (2001) analyzes optimal policy rules under different changes in specification of the Rudebusch-Svensson model. In this paper, we find minimax, not optimal, policy rules corresponding to different specifications of uncertainty. Further, the uncertainty we consider is more general than in Rudebusch (2001). Onatski and Stock (2002) analyze stability robustness of Taylor-type policy rules in the Rudebusch-Svensson model. Here we extend their analysis in a number of important ways. First, we study robust performance

[5] Due to technical constraints spelled out in the paper, we analyze only uncertainty about the Phillips curve under the SM approach. Neither the real-time data uncertainty, nor uncertainty about the IS curve are considered.

of the policy rules. We compute optimally robust policy rules and not just sets of the rules not destabilizing the economy that were computed in Onatski and Stock. Moreover, our more precise analysis makes it possible to study combined model, real-time data, and shock uncertainty in one structure. Second, we carefully calibrate the magnitude of the uncertainty about the Rudebusch-Svensson model using Bayesian estimation techniques.

To our knowledge, there are no previous attempts to estimate economic models with a goal of minimizing model ambiguity. This is essentially what we do in the part of the paper concerned with the Set Membership identification. Chamberlain (2000) uses max-min expected utility theory in estimation, but considers parametric estimation with a given set of possible models. A main task of our analysis is to construct the empirically plausible model set. Other broadly related papers use the theory of information based complexity, which is a general theory from which SM is derived. Information based complexity theory has been used and extended in economics by Rust (1997) and Rust, Traub, and Wozniakowski (2002) to study the numerical solution of dynamic programming problems.

In the next section of the paper we describe formulations of models of model uncertainty at a formal level, and show that parametric and shock uncertainty must be considered separately. Section 3 applies the Model Error Modeling approach to analyze robust monetary policies under uncertainty about Rudebusch-Svensson model. Section 4 models the Phillips curve and uncertainty associated with it using the Set Membership identification approach. Section 5 concludes.

2 Consequences of Different Uncertainty Models

2.1 Overview

The general issue that we consider in this paper is decision making under uncertainty. In particular, we focus on the policy-relevant problem of choosing interest rate rules when the true model of the economy is unknown and may be subject to different sources of uncertainty. The goal of the paper is to provide characterizations of the empirically relevant sources of uncertainty, and to design policy rules which account for that uncertainty.

As we noted in the introduction, we distinguish between different sources of uncertainty, each of which can lead to different implications for policy rules. To be concrete, it helps to introduce a standard state-space model:

$$\begin{aligned} x_{t+1} &= Ax_t + Bi_t + C\varepsilon_{t+1} \\ y_t &= Dx_t + Gi_t + F\eta_t \\ z_t &= Hx_t + Ki_t \end{aligned} \tag{1}$$

Here x_t is a vector state variables, i_t is a vector of controls, ε_t is a driving shock vector, y_t is a vector of observations, η_t is an observation noise and z_t is a vector of target variables. In the standard full information case, $F = 0$. By *model uncertainty* we mean uncertainty both about the parameters of the model and the model specification. In (1) this means variations in the A and B matrices (parametric uncertainty) as well as the makeup of the x_t vector (perhaps it omits relevant variables). By *shock uncertainty* we mean uncertainty about the serial correlation properties of the noise ε_t. By *data uncertainty* we mean uncertainty about the real-time data generating process. We capture this uncertainty by allowing variation in matrices D and G and specifying a range of possible spectral characteristics of η_t. Each of these sources of uncertainty enters the model (1) in a different way and can therefore lead to different outcomes.

One approach to model uncertainty which has been widely used was developed by Hansen and Sargent (2002). They absorb all sources of uncertainty into an additional shock to the system, and therefore collapse all uncertainty into shock uncertainty. Focusing on the full information case, they suppose that the first line of (1) is replaced by:

$$x_{t+1} = Ax_t + Bi_t + C(\varepsilon_{t+1} + w_{t+1}) \tag{2}$$

The decision maker then designs a control strategy that will be insensitive to sequences of w_t of a reasonable size (described in more detail below). While this approach seems quite general and unrestrictive, in the next section we develop a simple example which shows that the other sources of uncertainty cannot be collapsed to shock uncertainty.

To gain some intuition for the results, suppose that the decision maker uses a feedback control rule $i_t = kx_t$, and for simplicity assume C is the identity matrix. Then (2) becomes:

$$x_{t+1} = (A - Bk)x_t + \varepsilon_{t+1} + w_{t+1}.$$

Next, suppose that the true data generating process was a purely parametric perturbation of (1), so that A is replaced by $A + \hat{A}$, and B by $B + \hat{B}$. In this case we have that the model uncertainty perturbation is $w_{t+1} = (\hat{A} - \hat{B}k)x_t$, which clearly varies with the control rule k. Therefore in this setting it is impossible to evaluate the model uncertainty in the absence of the control rule. Different control rules will lead to different magnitudes of uncertainty, which cannot be judged ex-ante. This intuition is made more precise in the following section.

2.2 A Simple Example

In this section we build an example showing that two reasonable views of model uncertainty in a small macroeconometric model of the US economy may lead to drastically different policy implications. We consider a two-equation purely backward-looking model of the economy proposed and estimated by Rudebusch and Svensson (1999). This model will be the benchmark for the rest of the paper as well, and is given by:

$$\begin{aligned} \pi_{t+1} &= \underset{(.08)}{.70}\pi_t - \underset{(.10)}{.10}\pi_{t-1} + \underset{(.10)}{.28}\pi_{t-2} + \underset{(.08)}{.12}\pi_{t-3} + \underset{(.03)}{.14}y_t + \varepsilon_{\pi,t+1} \\ y_{t+1} &= \underset{(.08)}{1.16}y_t - \underset{(.08)}{.25}y_{t-1} - \underset{(.03)}{.10}(\bar{\imath}_t - \bar{\pi}_t) + \varepsilon_{y,t+1} \end{aligned} \tag{3}$$

The standard errors of the parameter estimates are given in parentheses. Here the variable y stands for the gap between output and potential output, π is inflation and i is the federal funds rate. All the variables are quarterly, measured in percentage points at an annual rate and demeaned prior to estimation, so there are no constants in the equations. Variables $\bar{\pi}$ and $\bar{i}$ stand for four-quarter averages of inflation and federal funds rate respectively.

The first equation is a simple version of the Phillips curve. The coefficients on the lags of inflation in the right hand side of the equation sum to one, so that the Phillips curve is vertical in the long run. The second equation is a variant of the IS curve. A policy maker can control the federal funds rate and wants to do so in order to keep y and π close to their target values (zero in this case).

In general, the policy maker's control may take the form of a contingency plan for her future settings of the federal funds rate. We, however, will restrict attention to the Taylor-type rules for the interest rate. As emphasized by McCallum (1988) and Taylor (1993), simple rules have the advantage of being easy for policymakers to follow and also of being easy to interpret. In this section, we will assume that the policy maker would like to choose among the following rules:

$$i_t = g_\pi \bar{\pi}_{t-1} + g_y y_{t-2} \tag{4}$$

Here, the interest rate reacts to both inflation and the output gap with delay. The delay in the reaction to the output gap is longer than that in the reaction to the inflation because it takes more time to accurately estimate the gap.

Following Rudebusch and Svensson, we assume that a policy maker has the quadratic loss:[6]

$$L_t = \bar{\pi}_t^2 + y_t^2 + \frac{1}{2}(i_t - i_{t-1})^2.$$

Had the policy maker been sure that the model is correctly specified, she could have used standard frequency-domain methods to estimate expected loss for any given policy rule (4). Then she could find the optimal rule numerically. We will however assume that the policy maker has some doubts about the model. She wants therefore to design her control so that it works well for reasonable deviations from the original specification.

One of the most straightforward ways to represent her doubts is to assume that the model parameters may deviate from their point estimates as, for example, is assumed in Brainard (1967). It is also likely, that the policy maker would not rule out misspecifications of the model's lag structure. As Blinder (1997) states, "Failure to take proper account of lags is, I believe, one of the main sources of central bank error."

For the sake of illustration, we will assume that the policy maker is willing to contemplate a possibility that one extra lag of the output gap in the Phillips curve and IS equations and one extra lag of the real interest rate in the IS equation were wrongfully omitted in the original model. She therefore re-estimates Rudebusch-Svensson model with the additional lags. The re-estimated model has the following form:

$$\begin{aligned} \pi_{t+1} &= \underset{(.08)}{.70}\pi_t - \underset{(.10)}{.10}\pi_{t-1} + \underset{(.10)}{.28}\pi_{t-2} + \underset{(.09)}{.12}\pi_{t-3} + \underset{(.10)}{.14}y_t + \underset{(.10)}{.00}y_{t-1} + \varepsilon_{\pi,t+1} \\ y_{t+1} &= \underset{(.08)}{1.13}y_t - \underset{(.12)}{.08}y_{t-1} - \underset{(.08)}{.14}y_{t-2} - \underset{(.14)}{.32}(\bar{\imath}_t - \bar{\pi}_t) + \underset{(.14)}{.24}(\bar{\imath}_{t-1} - \bar{\pi}_{t-1}) + \varepsilon_{y,t+1} \end{aligned} \tag{5}$$

[6]Inclusion of the term $(i_t - i_{t-1})^2$ in the loss function is somewhat controversial. Our results will not depend on whether this term is included in the loss function or not.

Then she obtains the covariance matrix of the above point estimates and tries to design her control so that it works well for reasonable deviations of the parameters from the point estimates. For example, she may consider all parameter values inside the 90% confidence ellipsoid around the point estimates.[7]

Technically, the above control design problem may be hard (because of the large number of parameters), but conceptually, it is straightforward. We will soon return to this problem, but for now let us assume that the policy maker decides to get around the possible technical difficulties. She therefore considers a different, more manageable, representation of the model uncertainty. Therefore she applies the Hansen and Sargent (2002) approach described above. For our example, the elements of (1) are $x_t = (\pi_t, \pi_{t-1}, \pi_{t-2}, \pi_{t-3}, y_t, y_{t-1}, i_{t-1}, i_{t-2}, i_{t-3})'$, $y_t = x_t$ (the state of the economy is perfectly observable), $\varepsilon_t = \left(\frac{\varepsilon_{\pi t}}{\sqrt{Var\varepsilon_{\pi t}}}, \frac{\varepsilon_{yt}}{\sqrt{Var\varepsilon_{yt}}}\right)'$, and $z_t = \left(\pi_t, y_t, \frac{i_t - i_{t-1}}{\sqrt{2}}\right)'$.

As we note above, under this approach the policy maker wants to choose rules which are insensitive to the distortion w_t in (2). More precisely, the policy maker solves the following problem:

$$\min_{\mathcal{F}} \max_{\{w_t\}} E \sum_{t=0}^{\infty} \beta^t z_t' z_t \quad s.t.\ E \sum_{t=0}^{\infty} \beta^t w_t' w_t < \eta. \tag{6}$$

where minimum is taken over all policies from the admissible set $\mathcal{F}$ (set of policy rules (4) in our example), parameter η regulates the size of the model uncertainty, and β is a discount factor. Here we consider the case $\beta \to 1$.

Problem (6) can be stated in an equivalent form

$$\min_{\mathcal{F}} \max_{\{w_t\}} E \sum_{t=0}^{\infty} \left(\beta^t z_t' z_t - \theta \beta^t w_t' w_t\right) \tag{7}$$

where θ is the Lagrange multiplier on the constraint in (6). A smaller value of θ corresponds to stronger preference for robustness as described in Hansen and Sargent (2002). Giordani and Söderlind (2002) explain how to solve this problem numerically for any given value of θ.

Anderson, Hansen, and Sargent (2000) propose to discipline the choice of θ by computing Bayesian probability of error in distinguishing between reference model (1) and distorted model (2), both controlled by the robust rule corresponding to θ. If the detection error probability is large, say larger than 0.1, the models are difficult to statistically distinguish from each other and the robustness parameter θ is said to correspond to a reasonable degree of model uncertainty. At the moment we will not bother to calibrate θ using the detection error probabilities. Instead, we will take θ to be the smallest robustness parameter such that there exists a rule from $\mathcal{F}$ corresponding to a finite value of problem (7). In other words, we are looking for a Taylor-type policy rule that "works best" when the size of the model uncertainty, η, tends to infinity.

[7] In the later sections of the paper we will discuss a more systematic way of representing and estimating the model uncertainty.

It can be shown that the robust policy rule corresponding to the "smallest possible" θ minimizes the so-called H_∞ norm of the closed loop system transforming noise ε into the target variables z (see Hansen and Sargent (2002)). It is therefore easy to find such an "extremely robust" rule numerically using, for example, commercially available Matlab codes to compute the H_∞ norm. Our computations give the following rule:

$$i_t = 3.10\bar{\pi}_{t-1} + 1.41y_{t-2}. \tag{8}$$

Now let us return to our initial formulation of the problem. Recall that originally we wanted to find a policy rule that works well for all deviations of the parameters of the re-estimated model (5) inside a 90% confidence ellipsoid around the point estimates. Somewhat surprisingly, the above "extremely robust" rule does not satisfy our original criterion for robustness. In fact, it destabilizes the economy for deviations from the parameters' point estimates inside as small as a 20% confidence ellipsoid. More precisely, policy rule (8) results in infinite expected loss for the following perturbation of the Rudebusch-Svensson (RS) model:

$$\begin{aligned}
\pi_{t+1} &= .68\pi_t - .13\pi_{t-1} + .35\pi_{t-2} + .10\pi_{t-3} + .30y_t - .15y_{t-1} + \varepsilon_{\pi,t+1} \quad (9)\\
y_{t+1} &= 1.15y_t - .07y_{t-1} - .18y_{t-2} - .51\left(\bar{\imath}_t - \bar{\pi}_t\right) + .41\left(\bar{\imath}_{t-1} - \bar{\pi}_{t-1}\right) + \varepsilon_{y,t+1}.
\end{aligned}$$

Let us denote independent coefficients[8] of the above model, the re-estimated the RS model (5), and the original RS model as c, c_1, and c_0 respectively. That is

$$\begin{aligned}
c &= (-.13, .35, .10, .30, -.15, 1.15, -.07, -.18, -.51, .41)',\\
c_1 &= (-.10, .28, .12, .14, .00, 1.13, -.08, -.14, -.32, .24)',\\
c_0 &= (-.10, .28, .12, .14, 0, 1.16, -.25, 0, -.10, 0)'.
\end{aligned}$$

And denote the covariance matrix of the coefficients in the re-estimated model (5) as V. Then we have

$$\begin{aligned}
(c - c_1)'V^{-1}(c - c_1) &= 6.15\\
(c_0 - c_1)'V^{-1}(c_0 - c_1) &= 5.34.
\end{aligned}$$

Both numbers are smaller than the 20% critical value of the chi-squared distribution with 10 degrees of freedom. This may be interpreted as saying that both the original Rudebusch-Svensson model and the perturbed model are statistically close to the encompassing re-estimated model.

Why does our "extremely robust" rule perform so poorly? It is not because other rules work even worse. For example, we numerically checked that the famous Taylor rule $i_t = 1.5\bar{\pi}_{t-1} + 0.5y_{t-2}$ guarantees stability of the economy at least for all deviations inside 60% confidence ellipsoid. Rule (8) works so bad simply because it was not designed to work

[8] Recall that the sum of coefficients on inflation in the Phillips curve is restricted to be equal to 1. We therefore exclude the coefficient on the first lag of inflation from the vector of independent coefficients.

well in such a situation. To see this, note that our original description of the model uncertainty allowed deviations of the slope of the IS curve from its point estimate. Therefore our ignorance about this parameter will be particularly influential if we were to use a very aggressive control rule. It may even be consistent with instability under such an aggressive rule. No effects of this kind are allowed under the Hansen and Sargent's "control friendly" description of the model uncertainty. Under this description, ignorance is equally influential for any policy rules. The specific interaction between aggressiveness of policy rules and uncertainty about slope of the IS curve is not taken into account. This lack of structure in the uncertainty description turns out to be dangerous because the resulting robust rule happens to be quite aggressive.

The example just considered should not be interpreted in favor of a particular description of uncertainty. Its message should rather be understood as follows. When designing robust policy rules we must carefully specify and thoroughly understand the model uncertainty that we are trying to deal with. Robust policy rules may be fragile with respect to reasonable changes in the model uncertainty specification. In the next sections therefore we try to rigorously approach the specification of the model uncertainty. We first use an approach based on model error modeling to estimate the uncertainty about the Rudebusch-Svensson model introduced above. Later we weaken our assumptions, and consider a set membership approach to joint model estimation and uncertainty evaluation.

3 Bayesian Model Error Modeling

Looking ahead, we will use frequency domain methods to evaluate uncertainty robustness of policy rules, so it is useful to reformulate our general setting of Section 2 using the language of lag polynomials. Suppose that a reference model of the economy has the following ARX form:

$$x_t = A_0(L)x_{t-1} + B_0(L)i_t + \varepsilon_t \tag{10}$$

where x_t is a vector of economic variables, i_t is a policy instrument, ε_t are white noise shocks, and the coefficients of matrix lag polynomials $A_0(L)$ and $B_0(L)$ satisfy a number of constraints reflecting specific features of the reference model. Note that the majority of purely backward-looking models of the economy can be represented in the above reduced form. In particular, the Rudebusch-Svensson model that we work with in the numerical part of this section can be viewed as a constrained ARX. An important question of how to choose a reference model of the economy is not considered until the next section of the paper.

Suppose further that a reference model of the observable economic data, y_t, has the form of the following noisy linear filter of the process x_t:

$$y_t = D_0(L)x_t + \eta_t.$$

For example, the reference model of the observable data may describe the real time data as n-period lagged actual data, that is $y_t = x_{t-n}$. In this case the white noise η_t has zero variance. We assume that policy makers use observed data to form a policy rule from an admissible class:

$$i_t = f(y_t, y_{t-1}, ..., i_{t-1}, i_{t-2}, ...)$$

For example, a possible class of admissible rules may consist of all linear rules $i_t = k \cdot y_t$.

Model (10) can be estimated using, say, OLS. The obtained estimate can then be used to formulate the optimal policy rule from the admissible class, where the criterion of optimality is based on the quadratic loss function:

$$L_t = \left|\left|\Lambda(L)(x_t, i_t)'\right|\right|^2.$$

where the norm $||\cdot||$ is the standard Euclidean length of a vector. The weighting matrix Λ may in general depend on the lag operator L, so that the loss depends on lagged variables.

The quality of the policy rule obtained in the above way will depend on the accuracy of the reference models for the economic dynamics and the observable data. In general, these models will not be accurate. We assume that the true models of the economy and the real time data encompass the reference models as follows:

$$x_t = (A_0(L) + A(L))\, x_{t-1} + (B_0(L) + B(L))i_t + (1 + C(L))\varepsilon_t \quad (11)$$

$$y_t = (D_0(L) + D(L))\, x_t + (1 + F(L))\eta_t \quad (12)$$

Here $A, B, C, D,$ and F are relatively unconstrained matrix lag polynomials, and η_t is the fundamental error of the process governing generation of observable data in the real time.

Let us assume that the central bank wants to design a policy rule that works well not only for the reference models but also for statistically plausible deviations from the reference models having the form (11) and (12). The set of deviations will be defined by a set of restrictions on the matrix lag polynomials $A, B, C, D,$ and F. The restrictions on $A(L)$ and $B(L)$ will describe the economic model uncertainty. This uncertainty includes econometric uncertainty about point estimates of $A_0(L)$ and $B_0(L)$, and uncertainty about restrictions on the coefficients of $A_0(L)$ and $B_0(L)$ introduced by the reference model. The restrictions on $C(L)$ will describe economic shock uncertainty. Models with non-zero $C(L)$ characterize economic shock processes as non-white stochastic processes. Finally, the restrictions on $D(L)$ and $F(L)$ will describe uncertainty about real time data generating process.

We define the restrictions on $A, B, C, D,$ and F as follows. For the i, j-th entry of the matrix polynomial $A(L)$ let:

$$A_{ij}(L) \in \left\{W_{A_{ij}}(L)\Delta_{A_{ij}}(L) : \left|\left|\Delta_{A_{ij}}(L)\right|\right|_\infty \leq 1\right\}$$

where $W_{A_{ij}}(L)$ is a scalar lag polynomial weigh and $\Delta_{A_{ij}}(L)$ is an uncertain scalar linear filter with gain less than or equal to unity. For the polynomials $B, C, D,$ and F the restrictions are defined similarly, that is, for example:

$$B_{ij}(L) \in \left\{W_{B_{ij}}(L)\Delta_{B_{ij}}(L) : \left|\left|\Delta_{B_{ij}}(L)\right|\right|_\infty \leq 1\right\}.$$

If all weights $W_{(\cdot)ij}(L)$ are zero, then the set of plausible deviations from the reference models is empty and the central bank pays attention only to the reference models. In general, the spectral characteristics of the weighting lag polynomials may be chosen so that the resulting set of models includes relatively more plausible deviations from the reference models. The question of how to choose the weights in practice will be discussed below.

Our final goal is to find a policy rule to minimize the loss for the worst possible model (11) and (12) subject to restrictions defined above. As explained in Onatski (2001), it is possible to find an upper bound on such a worst possible loss numerically. We will therefore look for the policy rules that minimize the upper bound on the worst possible loss. Unfortunately, there are no theoretical guarantees that the upper bound is tight. However, the bound has an appealing interpretation of the exact worst possible loss under slowly time-varying uncertainty (see Paganini (1996)). Numerical exercises with the Rudebusch-Svensson model suggest that the bound may be very tight for relatively simple dynamic models and uncertainty characterizations.

3.1 Bayesian Calibration of Uncertainty

How should we choose weights $W_{(\cdot)ij}(L)$ so that the corresponding restrictions on A, B, C, D, and F describe a statistically plausible model uncertainty set? In this section we approach this question using the Model Error Modeling (MEM) methodology recently developed in the system identification literature (see Ljung (1999)).

Let us denote the true errors of the reference models for economic dynamics and the real time data as e_t and e_t^d. From (11) and (12) we have:

$$e_t = A(L)x_{t-1} + B(L)i_t + (1 + C(L))\varepsilon_t \quad (13)$$

$$e_t^d = D(L)x_t + (1 + F(L))\eta_t. \quad (14)$$

That is, the true models for the real time data and the economic dynamics imply the above models for the errors of the reference models.

Knowing the reference model parameters $A_0(L), B_0(L)$, and $D_0(L)$, we can obtain the data on e_t, e_t^d from the real time data y_t and the (final) data on x_t and i_t. Then we can estimate model error models (13) and (14). As we mentioned before, the matrix lag polynomials in the true model should be relatively unrestricted. However, some restrictions may still be plausible *a priori*. For example, we might want to see the coefficients of the lag polynomials A, B, C, D and F exponentially decaying because we believe that the described economic and real-time data generating processes have relatively short memory.

To incorporate some *a priori* restrictions on the parameters of the model error models, we will estimate the models using Bayesian methods. After the estimation the whole posterior distribution of the parameters will be available. In particular, we will obtain the posterior distribution of the coefficients of $A_{ij}(L)$, $B_{ij}(L)$, ..., $F_{ij}(L)$. Therefore, we will know the posterior distribution of the filter gains $|A_{ij}(e^{i\omega})|$, $|B_{ij}(e^{i\omega})|$, ..., $|F_{ij}(e^{i\omega})|$ at each frequency point ω.

Let us define $a_{ij}(\omega), b_{ij}(\omega), ..., f_{ij}(\omega)$ so that

$$\begin{aligned} \Pr\left(\left|A_{ij}(e^{i\omega})\right| \leq a_{ij}(\omega)|data\right) &= 95\% \\ &... \\ \Pr\left(\left|F_{ij}(e^{i\omega})\right| \leq f_{ij}(\omega)|data\right) &= 95\% \end{aligned}$$

We can then choose an ARMA weight $W_{A_{ij}}(L)$ so that its gain $\left|W_{A_{ij}}(e^{i\omega})\right|$ approximates $a_{ij}(\omega)$ well at each frequency point. We can choose the other weights $W_{(\cdot)_{ij}}(L)$ similarly.

The described procedure of choosing weights still leaves a possibility that some statistically implausible models will be included in the model uncertainty set. This happens because we define the weights frequency-by-frequency so that the weights are essentially a 95% envelope of gains of different filters drawn from posterior distribution for $A_{ij}(L)$, $B_{ij}(L)$, ..., $F_{ij}(L)$. Such an envelope may be very different from the sampled gain functions.

3.2 A Numerical Example

In this section we analyze robustness of Taylor-type policy rules for model, data, and shock uncertainty about the Rudebusch-Svensson (RS) model given in (3) above. We also consider the quadratic loss studied by Rudebusch and Svensson:

$$L_t = \bar{\pi}_t^2 + y_t^2 + \frac{1}{2}(i_t - i_{t-1}).$$

The real-time data on the output gap and inflation was kindly provided to us by Athanasios Orphanides. The data is relatively short. It starts from 1987:1 and ends at 1993:04. In this sample, the lagged data turns out to be a better predictor of the real time data than the current actual (final) data. We therefore use as our reference model for the real-time generating process the following:

$$\begin{aligned} \pi_t^* &= \pi_{t-1}, \\ y_t^* &= y_{t-1}. \end{aligned}$$

That is, our reference assumption is that the real time data on inflation and the output gap are equal to the lagged data on inflation and the output gap.

Using the Rudebusch-Svensson data set kindly provided to us (some time ago) by Glenn Rudebusch, we compute the errors of the RS Phillips curve, e_{t+1}^{π}, and the IS curve, e_{t+1}^{y}. Using Athanasios Orphanides' data, we compute the errors of our reference model for the real time data on inflation, $e_t^{\pi,data}$, and the output gap $e_t^{y,data}$. We then model the reference

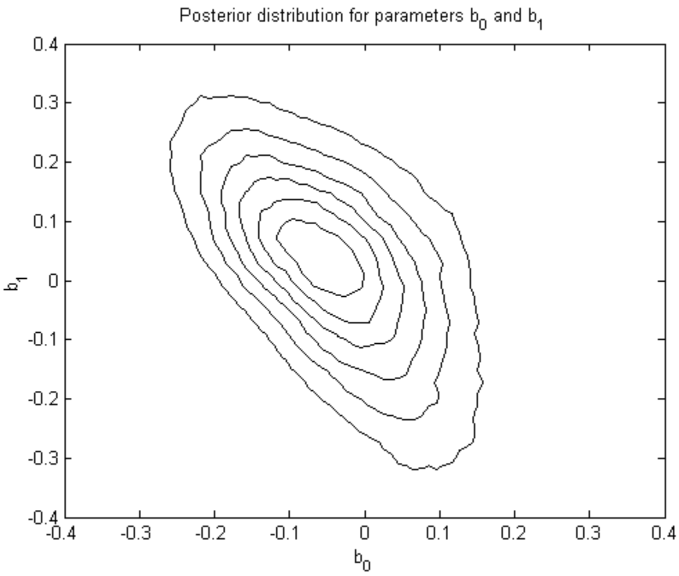

Figure 1: Posterior distribution for the parameters b_0 and b_1.

models' errors as follows:[9]

$$
\begin{aligned}
e^{\pi}_{t+1} &= \sum_{i=0}^{4} a_i L^i(\pi_t - \pi_{t-1}) + \sum_{i=0}^{4} b_i L^i y_t + \left(1 + \sum_{i=1}^{4} c_i L^i\right) \varepsilon^{\pi}_{t+1} \\
e^{y}_{t+1} &= \sum_{i=0}^{4} d_i L^i y_t + \sum_{i=0}^{4} f_i L^i \left(i_t - \pi_t\right) + \left(1 + \sum_{i=1}^{4} g_i L^i\right) \varepsilon^{y}_{t+1} \\
e^{\pi,data}_{t} &= \sum_{i=0}^{3} h_i L^i \pi_t + \left(1 + \sum_{i=1}^{2} k_i L^i\right) \eta^{\pi}_t \\
e^{y,data}_{t} &= \sum_{i=0}^{3} l_i L^i y_t + \left(1 + \sum_{i=1}^{2} m_i L^i\right) \eta^{y}_t
\end{aligned}
$$

Assuming diffuse priors for all parameters, we sampled from the posterior distributions of coefficients $a, b, c, ..., m$ and the posterior distributions of the shock variances using the algorithm of Chib and Greenberg (1994) based on Markov Chain Monte Carlo simulations. We obtained six thousand draws from the distribution and dropped the first thousand draws.

In Figure 1 we provide a contour plot of the estimate of the joint posterior density of the parameters b_0 and b_1. These parameters can roughly be interpreted as measuring error of the RS models' estimates of the effect of the output gap on inflation. The picture demonstrates that the Rudebusch-Svensson model does fairly good job in assessing the size of the effect of a one time change in the output gap on inflation. However, there exist some chances that

[9] Our model for the real-time data errors has fewer lags than that for the Rudebusch-Svensson model errors because our sample of the real-time data is much shorter than our sample of the final data

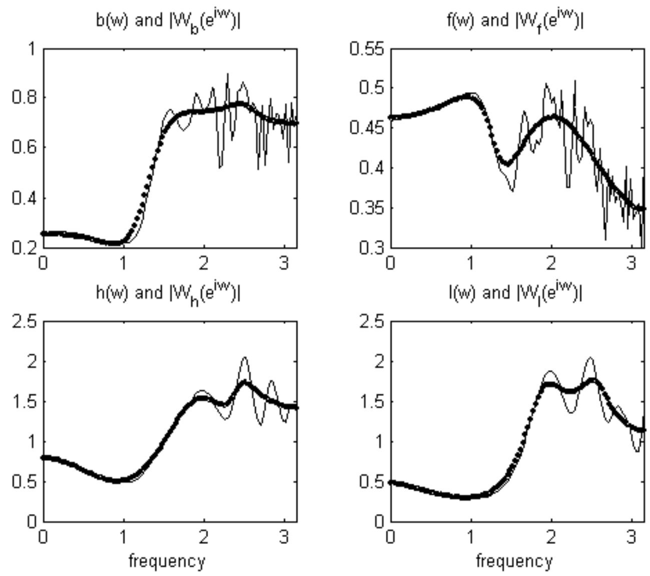

Figure 2: 95% envelopes of the sampled gains of different uncertainty channels. Dotted lines correspond to rational approximations to the envelopes.

the effect is either more spread out over the time or, vice versa, that the initial response of inflation overshoots its long run level.

Using our sample of the posterior distribution for parameters $a, b, ...,$ and m we simulated posterior distributions of $\sum_{t=0}^{4} a_t e^{it\omega}, \sum_{t=0}^{4} b_t e^{it\omega}, ...,$ and $1 + \sum_{t=1}^{2} m_t e^{it\omega}$ on a grid over the unit circle in the complex plain. For each frequency in the grid, we found thresholds $a(\omega), b(\omega), ..., m(\omega)$ such that 95% of the draws from the above posterior distribution had absolute value less than the thresholds.[10] We then define weights $W_a(L), W_b(L), ...,$ and $W_m(L)$ as rational functions of L with numerators and denominators of fifth degree whose absolute values approximate $a(\omega), b(\omega), ..., m(\omega)$ well in the sense of maximizing the least squares fit on our grid in the frequency domain.

In Figure 2 we plot the thresholds $b(\omega), f(\omega), h(\omega)$, and $l(\omega)$ together with the absolute values of the corresponding weights. For the three out of the four reported thresholds more uncertainty is concentrated at high frequencies. However, for $f(\omega)$ which corresponds to uncertainty about effect of the real interest rate on the output gap, there exists a lot of uncertainty at low frequencies.

We computed an upper bound for the worst possible loss of the Taylor rule:

$$i_t = 1.5\bar{\pi}_t^* + 0.5y_t^*$$

[10] We define $k(\omega)$ and $m(\omega)$ as the numbers such that 95% of the draws from the posterior for $1+\sum_{t=1}^{2} k_t e^{it\omega}$ and $1 + \sum_{t=1}^{2} m_t e^{it\omega}$ lie closer to the posterior mean of $1 + \sum_{t=1}^{2} k_t e^{it\omega}$ and $1 + \sum_{t=1}^{2} m_t e^{it\omega}$ than $k(\omega)$ and $m(\omega)$ respectively. As Orphanides (2001) notes, the noise in real time data may be highly serially correlated. By centering the real time data uncertainty description around the posterior mean for k and m (instead of zero) we essentially improve our reference model for the real time data generating process.

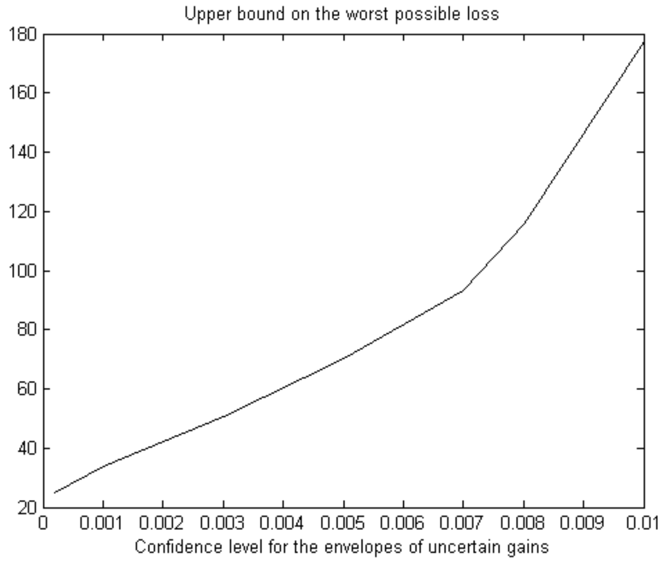

Figure 3: Upper bound on the worst possible loss for the Taylor rule.

where $\bar{\pi}_t^*$ is a four quarter average of the real-time data on inflation and y_t^* is the real time data on the output gap. This policy rule resulted in infinite loss. Our description of uncertainty turns out to include some models that are unstable under the Taylor rule.

We then analyzed smaller perturbations to the reference model. We redefine $a(\omega), ..., m(\omega)$ as the envelopes for the posterior distribution of uncertain gains corresponding to lower than 95% confidence level. The graph of the upper bound on the worst possible loss for the Taylor rule for different confidence levels is shown in Figure 3. We see that the worst possible loss becomes ten times higher than the loss under no uncertainty for the confidence levels as small as 1%! This roughly means that even among the 50 draws from the posterior distribution which are closest to zero (out of 5000 draws) we can find such parameters $a, b, ..., m$ that correspond to an economic disaster if the Taylor rule were mechanically followed by the central bank.

Recall that according to our classification, the posterior distributions for $a, b, d,$ and f describe pure model uncertainty, the posterior distributions for c and g describe shock uncertainty, and the posterior distribution for the rest of the parameters describe real time data uncertainty. How do different kinds of uncertainty contribute to the miserable worst scenario performance of the Taylor rule? To answer this question, we increased the size of the specific uncertainty, keeping the coefficients corresponding to all other sources of uncertainty equal to the values closest to zero that were ever drawn in our simulations. The graph with contributions of different forms of uncertainty to the worst possible loss is given in Figure 4.

We see that the pure model uncertainty makes huge input into the worst possible scenario for the Taylor rule. This result supports the common believe that under the existing amount of model uncertainty, mechanically following any Taylor-type rule specification is an extremely dangerous adventure. Alternatively, it may be indicating that policy makers have

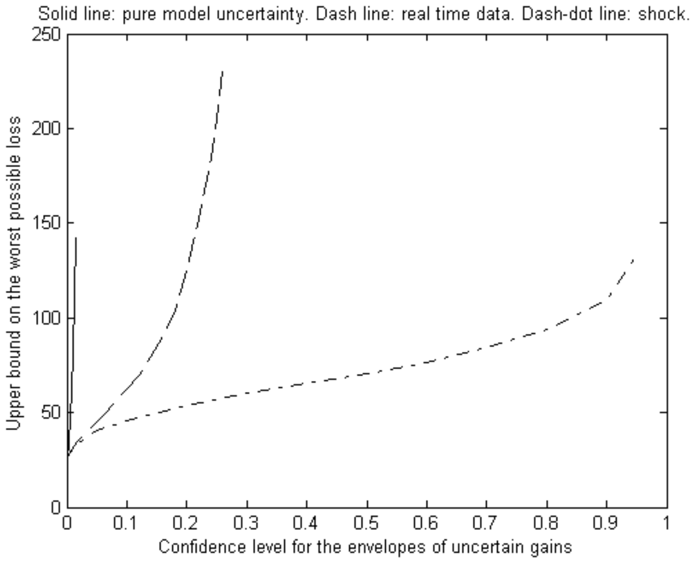

Figure 4: Upper bound on the worst possible loss. The solid line corresponds to pure model uncertainty, the dashed line corresponds to real time data uncertainty, and the dash-dot line corresponds to pure shock uncertainty.

strong priors about parameters $a, b, ..., m$ of the true model. The pure shock uncertainty is the least important source of uncertainty. Clearly, changing the spectral characteristics of the shock process cannot make the model unstable under the Taylor rule. Even for very broad confidence regions for the spectrum of the shock processes, the worst possible loss is only about 5 times larger than the loss under no uncertainty. The real time data uncertainty significantly contributes to the worst possible scenario for the Taylor rule. It is possible to design a linear filter procedure for generating real-time data from the actual data that will mimic the historical relationship between the real time data and the final data quite well and will lead to instability under the Taylor rule.

As we mentioned above, the poor performance of the Taylor rule under the uncertainty may be caused by our assumption of diffuse priors for the parameters of the true model. It may turn out that, contrary to our assumption, policy makers' priors are very informative. To see how informativeness of the priors affects the optimally robust Taylor-type rule, we do the following exercise. We start from a very informative prior for the parameters $a, b, ..., m$. The informative priors put a lot of weight on values of the parameters very close to zero. The optimally robust Taylor rule for such extremely informative prior is essentially the optimal Taylor-type rule under no uncertainty. Then we decrease informativeness of priors monotonically. We do this separately for the pure model, real data, and shock uncertainty parameters. We plot the corresponding optimally robust Taylor-type rules in Figure 5.

When the informativeness of the prior on the model uncertainty parameters falls, the optimally robust Taylor-type rules become less active in the inflation response and somewhat more active in the response to the output gap. This is in accordance with the results

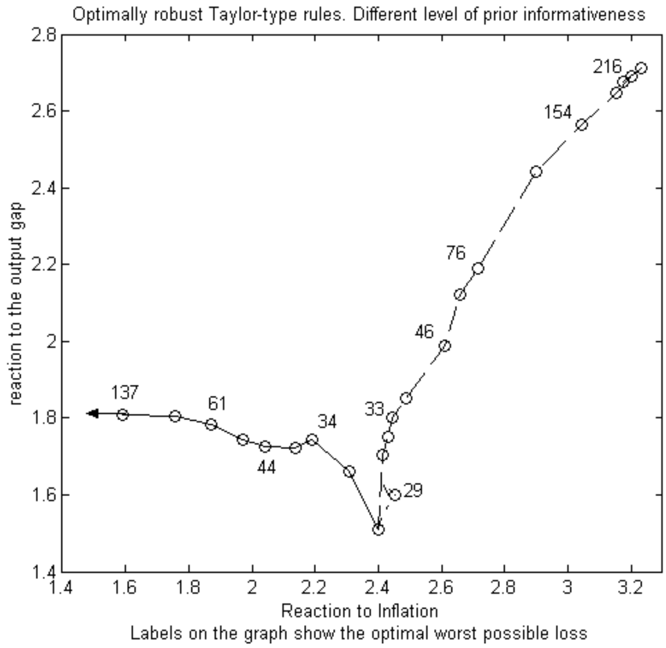

Figure 5: Optimally robust Taylor-type rules under different levels of prior informativeness. Labels on the graph show the worst possible losses corresponding to the robust rules. Solid line: pure model uncertainty. Dash line: pure real time data uncertainty. Dash-dot line: pure shock uncertainty.

of Onatski and Stock (2002). Contrary to our expectations, when the informativeness of the prior for the real time data uncertainty falls, the optimally robust rule become more aggressive both in its response to the output and in its response to inflation. This result is caused by the fact that we allow a possibility that the real-time data on inflation is systematically underestimating the actual inflation. In such a case, an active reaction on both changes in the real-time inflation data and changes in the real-time output gap data are needed to prevent inflation from permanently going out of control. Finally, the Taylor-type rules optimally robust against pure shock uncertainty are very close to the optimal Taylor-type rule under no uncertainty.

The level of the optimal worst possible loss (given on the graph) rises when the informativeness of priors falls. This happens because the size of the corresponding uncertainty increases. These levels supply a potentially useful piece of information on the plausibility of the corresponding prior. In our view, the levels of the worst possible loss less than or equal to about 80 correspond to plausible model uncertainty sets, and therefore to plausible priors on uncertainty parameters. Our computations show that the loss under no uncertainty for the benchmark Taylor rule is about 20. Therefore, the loss equal to 80 roughly corresponds to standard deviations for inflation and the output gap twice as large as those historically observed. Hence, it is not unreasonable to assume that the worst possible models implying losses about 80 must be considered seriously by policy makers.

Figure 6 shows a contour plot of the worst possible losses for a broad range of the Taylor-

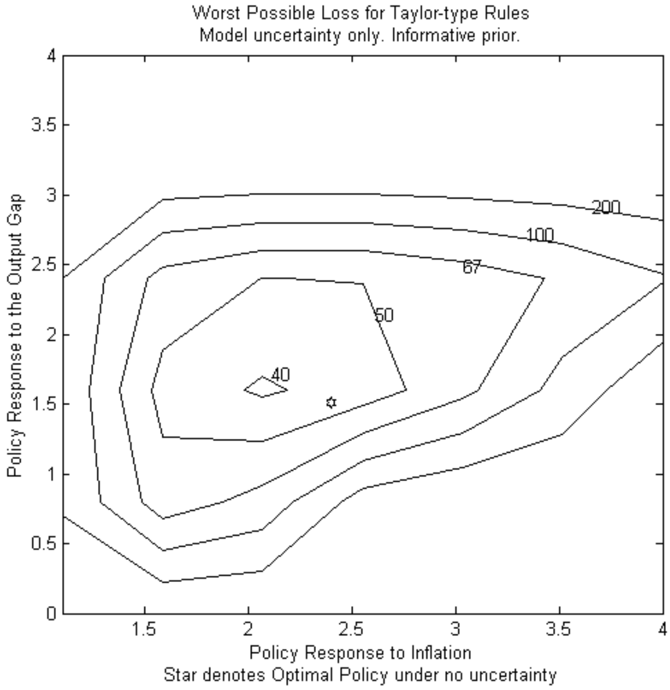

Figure 6: Worst possible loss for Taylor-type rules under model uncertainty only. Informative prior calibrated so that the worst possible loss under the Taylor rule is about 80.

type rules computed under an informative prior for pure model uncertainty. The level of informativeness of the prior was calibrated so that the worst possible loss under the Taylor rule was about 80. The optimally robust rule under such a prior is not very different from the optimal policy rule under no model uncertainty (denoted by a star at the graph). The level of robustness deteriorates quickly when one deviates from the optimally robust rule. We obtained qualitatively similar results, which we do not report here, for the real time data uncertainty and pure shock uncertainty.

So far, we combined uncertainty about the own dynamics of inflation in the Phillips curve, uncertainty about the dynamic effect of the output gap on inflation, uncertainty about the own dynamics of the output gap in the IS curve, and uncertainty about the dynamic effect of the real interest rate on the output gap into what we called the model uncertainty. Similarly, we combined uncertainty about news and noise in the error of the real-time data on the output gap and uncertainty about news and noise in the error of the real-time data on inflation into what we called the real-time data uncertainty.[11]

To get some insight on the relative importance of the different blocks of model uncertainty and the real-time data uncertainty, we computed the optimally robust Taylor-type rules

[11]Uncertainty about the coefficients l_i can be thought of as uncertainty about news in the error of the real-time data on the output gap. It is because the part of the real-time data error described with the help of l_i is correlated with the final data on the output gap. Similarly, uncertainty about the coefficients m_i can be thought of as uncertainty about noise in the error of the real-time data on the output gap.

Uncertainty Channel	Coefficient on Inflation	Coefficient on Output Gap	Worst Possible Loss
No uncertainty	2.4	1.5	13.1
Own dynamics of π	2.3	1.6	20.3
Effect of y on π	2.9	2.6	76.0
Own dynamics of y	2.2	2.0	29.6
Effect of r on y	2.1	1.7	146.1
News in the real-time output gap error	2.0	1.1	22.1
Noise in the real-time output gap error	2.3	1.4	16.1
News in the real-time inflation error	3.4	2.7	46.8
Noise in the real-time inflation error	2.3	1.5	16.2

Table 1: The coefficients of the robust optimal Taylor rules and corresponding worst possible losses for diffuse priors on different uncertainty channels.

corresponding to very informative (zero) prior on all channels of uncertainty except one, such as for example, uncertainty about the own dynamics of inflation in the Phillips curve. For each chosen channel of uncertainty we consider a relatively uninformative prior so that only this specific channel matters for the robust decision maker. Our results are reported in Table 1.

We see that the most dangerous block is represented by the uncertainty about the slope of the IS curve. The worst possible loss for the optimally robust rule corresponding to such uncertainty is an order of magnitude larger than the optimal loss under no uncertainty whatsoever. The least dangerous among uncertainty blocks representing model uncertainty is the uncertainty about the own dynamics of inflation. This result, however, is an artifact of our maintaining a vertical long-run Phillips curve. Had we allowed for a non-vertical Phillips curve in the long-run, the importance of uncertainty about the own dynamics of inflation would have been much higher.

Among all optimally robust rules reported in Table 1, only the rules corresponding to uncertainty about real-time data on the output gap are less aggressive than the optimal rule under no uncertainty. In fact, for the uncertainty about the coefficients l_i (which we interpret as uncertainty about news in the error of the real-time data on the output gap), the optimally robust Taylor-type rule has the coefficient on inflation 2, and the coefficient on the output gap 1.1. This is not far from the Taylor-type rule that best matches the Fed's historical behavior (see Rudebusch (2001)). On the contrary, the optimally robust rule corresponding to the uncertainty about news in the error of the real-time data on inflation is extremely aggressive. This finding supports our explanation (given above) of the fact that the combined real-time data uncertainty implies aggressive robust policy.

Uncertainty Channel	Coefficient on Inflation	Coefficient on Output Gap	Worst Possible Loss
No uncertainty	2.4	1.5	13.1
Own dynamics of π	2.3	1.5	17.2
Effect of y on π	2.2	1.6	20.8
Own dynamics of y	2.2	1.6	20.7
Effect of r on y	1.9	1.0	24.4
News in the real-time output gap error	2.1	1.2	17.8
Noise in the real-time output gap error	2.4	1.4	15.3
News in the real-time inflation error	2.4	1.8	21.2
Noise in the real-time inflation error	2.3	1.5	15.4

Table 2: The coefficients of the robust optimal Taylor rules and corresponding worst possible losses for diffuse priors on different uncertainty channels. Business cycle frequencies only.

We can further improve our analysis by focusing on specific frequency bands of the uncertainty. For example, we may be most interested in the uncertain effects of business cycle frequency movements in inflation and the output gap on the economy. Table 2 reports the coefficients of the optimally robust Taylor-type rules for different uncertainty blocks truncated so that the uncertainty is concentrated at frequencies corresponding to cycles with periods from 6 quarters to 32 quarters.[12] In contrast to Table 1, Table 2 does not contain very aggressive robust policy rules. Now most of the robust policy rules reported are less aggressive than the optimal rule under no uncertainty. As before, the most dangerous uncertainty is the uncertainty about the slope of the IS curve. Now however, the worst possible loss corresponding to this uncertainty is only 86% higher than the optimal loss under no uncertainty. Moreover, the corresponding optimally robust rule is the least aggressive of all the rules and pretty much consistent with the historical estimates of the Taylor-type rules.

The drastic change of results in Table 2 relative to the results reported in Table 1 is caused by the fact that the worst possible perturbations of the reference model are concentrated at very low frequencies. To see this, we computed the optimally robust Taylor rules for low frequency uncertainty (only cycles with periods longer than 32 quarters were allowed). The computed rules, reported in Table 3, are very aggressive and the corresponding worst possible losses are uniformly worse than those reported for business cycle frequencies uncertainty.[13]

[12]Technically, we multiply thresholds $a(\omega), b(\omega), ..., m(\omega)$ by zero for frequencies outside the range $[2\pi/32, 2\pi/6]$.

[13]The worst possible loss for the uncertainty about effect of y on π is larger than that reported in Table 1. A possible reason for this is that the rational weight approximating the threshold $b(\omega)$ multiplied by zero for $\omega > 2\pi/32$ is slightly larger (at very low frequencies) than the rational weight approximating $b(\omega)$ at all frequencies.

Uncertainty Channel	Coefficient on Inflation	Coefficient on Output Gap	Worst Possible Loss
No uncertainty	2.4	1.5	13.1
Own dynamics of π	2.5	1.7	18.3
Effect of y on π	3.6	3.0	82.7
Own dynamics of y	2.7	2.4	23.0
Effect of r on y	2.9	3.1	53.4
News in the real-time output gap error	2.3	1.6	19.6
Noise in the real-time output gap error	2.4	1.5	15.4
News in the real-time inflation error	3.8	2.9	37.8
Noise in the real-time inflation error	2.4	1.6	15.6

Table 3: The coefficients of the robust optimal Taylor rules and corresponding worst possible losses for diffuse priors on different uncertainty channels. Low frequencies only.

Since the optimally robust rules reported in Table 2 have relatively small worst possible losses, we might hope that the combined uncertainty (estimated with the diffused prior on all uncertainty blocks) truncated to the business cycles frequency is not too big to allow decent robust performance for at least some of the Taylor-type rules. Unfortunately, this is not so. Infinite worst possible losses result for all policy rules in the plausible range. Finite losses become possible only when we redefine $a(\omega), b(\omega), ..., m(\omega)$ so that the uncertainty is represented by less than 20% of the draws closest to zero from the posterior distribution. The optimally robust Taylor-type rule for the uncertainty corresponding to 10% of the "smallest" draws from the posterior distribution has the coefficient on inflation 1.4 and the coefficient on the output gap 0.7 which is surprisingly close to the Taylor (1993) rule.

4 Set Membership Estimation

In the previous section we analyzed the uncertainty associated with a given model, the Rudebusch-Svensson model. We then developed an approach for quantifying this uncertainty based on model error modeling, and used this quantification as an input into a robust control problem. In this section, we analyze how to estimate a nominal model as well. We want to build in fewer *a priori* assumptions than in the MEM approach, to use measures of model uncertainty as a basis for estimation, and to link the choice of model to our ultimate control objectives. In summary, our goal in this section is to use time series data to construct (1) a set of models which could have generated the data and a description of the "size" of this set, (2) a baseline or nominal model which serves as an estimate of the data generating process, (3) a control policy which is optimal (in some sense) and reflects the model uncertainty.

4.1 Set Membership and Information Based Complexity

In their formulation of robust control theory, Hansen and Sargent (2002) consider, "a decision maker who regards his model as a good approximation in the sense that he believes that the data will come from an unknown member of a set of unspecified models near his approximating model." However, they are silent on where the "good approximating model," which we will refer to as the nominal model, comes from. The same applies to the other applications of robustness in economics, and to our specifications above. The specification of the nominal model typically follows from theory or is a simple empirical model, and it is typically estimated using standard statistical methods. However if agents do not trust their models, why should they trust their statistical procedures, which make explicit or implicit assumptions that the model is correctly specified? In this section we address the question of how to use observed data in order to simultaneously identify the nominal model and a class of alternative models.

We use a completely deterministic approach to model identification and estimation, which is known as set membership (SM) identification theory. These methods are based on the principles of robust control, and provide bounds on the model set which can be naturally integrated into the choice of a robust decision rule. While the unified treatment of estimation and control is the main benefit of the approach, an important drawback is that existing methods in set membership theory are limited to single-input single-output models. Therefore in this section we focus exclusively on the Phillips curve, and so suppose that policymakers directly control the output gap, or that the IS equation linking interest rates and the output gap is known and not subject to shocks. While this is limiting, our estimation results are of interest in their own right, and the Phillips curve provides a simple laboratory for comparing different estimation and control procedures.

The basic concepts of set membership identification theory are derived from information based complexity (IBC), which is a general theory which considers the solutions of continuous problems with discrete observations and potentially noisy data (see Traub, Wasilkowski, and Wozniakowski (1988)). The most widely used formulations consider deterministic problems with a worst-case criterion. IBC has been used and extended in economics by Rust (1997) and Rust, Traub, and Wozniakowski (2002) in the context of numerical solutions of dynamic programming problems. Set membership theory can be viewed as an application and extension of information based complexity to consider the estimation of dynamic control processes with noisy observations. The general idea of SM is to consider a set of models that could have generated the data and satisfy a number of weak *a priori* restrictions and further to approximate this relatively complex set by a ball covering it in the model space. Then the center of the ball becomes the nominal model, and the radius of the ball provides an upper bound on the size of the set of relevant models. Of course, the approximation can be done in a number of different ways. Good approximation algorithms make the approximating ball as small as possible.

Following Milanese and Taragna (1999) and Milanese and Taragna (2001) we now describe the SM formulation of the estimation problem. As noted above, we focus on the Phillips curve. We suppose that policymakers control the output gap y_t directly in order to influence inflation π_t.[14] Inflation is typically estimated to have a unit root, as in the Rudebusch-Svensson model, and this has some theoretical support (at least in backward looking models such as ours). Since below we need to impose the counterparts of stationarity assumptions, we therefore focus on the *changes* of inflation $x_t = \Delta\pi_t$ as the output (and later deduce the implied responses of levels of inflation).

Let $\mathcal{S}$ be the set of all possible causal linear time-invariant systems linking the output gap to changes in inflation. We identify any $S \in \mathcal{S}$ with its impulse response $h^S = \{h_0^S, h_1^S, h_2^S, \ldots\}$. We suppose that the unknown true model of the Phillips curve is some $M^0 \in \mathcal{S}$. The true model is thus not (necessarily) finitely parameterized, but we suppose that it is known *a priori* to belong to a subset of all possible models. Below we identify this subset by restrictions on the impulse responses of the models. The observed time series are thus noise-corrupted samples of the true model:

$$x_t = \sum_{s=0}^{t} h_s^{M^0} y_{t-s} + e_t, \tag{15}$$

where e_t is an additive noise corrupting the exact linear relationship. Here, for simplicity, we suppose that the economy was in the steady state prior to date 0, so that $x_s = y_s = 0,\ s < 0$.

The classical statistical approach to estimation assumes that the true model M^0 belongs to a finitely parameterized class of models $M(p)$ where $p \in \mathbb{R}^n$, and further that e is a stochastic noise with known distribution. In this case the estimation problem reduces to determining the parameter vector p. For example, a common specification is an ARX(j,k) model of the form:

$$x_t = \sum_{s=1}^{j} a_i x_{t-i} + \sum_{j=0}^{k} b_i y_{t-i} + e_t. \tag{16}$$

Note that this can be reformulated as in (15), where now the impulse responses are parameterized by $p = [a_1, \ldots, a_s, b_0, \ldots, b_k]' \in \mathbb{R}^{j+k+1}$. Under the standard orthogonality condition on the errors e_t, the parameters can be consistently estimated by OLS.

Clearly, under the classical approach, the only model uncertainty is in the value of the parameters. However it is more likely to be the case that the true model does not belong to the parametric model class $M(p)$, in which case we also need to consider unmodeled dynamics represented by unknown causal linear time-invariant operator Δ. In this case, (15) can be modeled as:

$$x_t = \sum_{s=0}^{t} h_s^{M(p)} y_{t-s} + \Delta(y_t, y_{t-1}, \ldots) + e_t \tag{17}$$

In this section we use set membership methods in order to jointly identify a parametric nominal model of the system and to obtain bounds on the size of the unmodeled dynamics. Our methods here are completely deterministic and based on a worst-case criterion, just as

[14] In our discussion of policy rules below, we augment the model with an exact IS equation.

in our robust control formulations. Further, the norm of Δ is a direct input into robust control problems, so the approach here naturally links robust estimation and control. One goal of this section is to analyze parametric nominal models such as (16). The best ARX model will be that which comes with the smallest estimated size of the unmodeled dynamics Δ. However our focus is on a more general approach which directly estimates the impulse response of the output to the input. While economists working with VARs typically analyze the impulse responses of variables to shocks, note that here we deal with the impulse response of the output (changes in inflation) to the controlled input (the output gap). By estimating the impulse responses directly we allow for a range of dynamics which may be difficult to capture in an ARX model.

Let us rewrite (15) in the following form:

$$x = Yh^{M^0} + e \tag{18}$$

where x is a vector of the N available observations of the change in inflation, e is a $N \times 1$ vector of the shocks and Y is a matrix with N rows and an infinite number of columns composed of the observations of the output gap,:

$$Y = \begin{bmatrix} y_0 & 0 & \dots & 0 & 0 & \dots \\ y_1 & y_0 & \dots & 0 & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ y_{N-1} & y_{N-2} & \dots & y_0 & 0 & \dots \end{bmatrix}. \tag{19}$$

The key assumptions of set membership theory consist of (1) specifying a compact set H of models that is presumed to contain the true model M^0 and (2) specifying a bounded set B for the noise e. These assumptions are relatively minimal, and are necessary in order to identify the system. Assumption (2) is relatively straightforward, and is commonly known as "unknown but bounded" noise. We assume that at each date t, e_t is bounded in absolute value by a constant ε:

$$B(\varepsilon) = \left\{ e = [e_0, \dots, e_{N-1}]' \in \mathbb{R}^N : ||e||_\infty = \max_{0 \le k \le N-1} |e_k| < \varepsilon \right\}.$$

In particular, we will not make any correlation assumptions on the noise.[15] There are different options for the norm on B, with ℓ_1, ℓ_2, and ℓ_∞ being the most common. The ℓ_∞ bound that we use naturally matches with some specifications of robust control. Concerning assumption (1), recall that we identify a model with its impulse response. Therefore we assume that the impulse responses lie in the following infinite dimensional compact set:

$$H(L, \rho) = \left\{ h \in \mathbb{R}^\infty : |h_t| \le L\rho^t, \ \forall t \ge 0 \right\}.$$

The we associate with H all models in which the immediate effect of the input on the output is bounded by L, and the impulse responses decay geometrically with factor ρ.

[15] If correlation properties of the shocks are known *a priori* they can be imposed. Similarly, the observations could be weighted to reflect time-varying noise bounds. See Milanese and Taragna (2001).

Combining our assumptions on the noise and the system, we can derive the set of all models which could have generated the observed data. This is a fundamental set in set membership identification, and is known as the *feasible systems set* (FSS):

$$FSS(L, \rho, \varepsilon) = \left\{ S \in \mathcal{S} : h^S \in H(L, \rho),\ ||x - Yh^S||_\infty \leq \varepsilon \right\}. \tag{20}$$

Describing the FSS is the main task of set membership theory, because under the maintained assumptions it is guaranteed to contain the true system, and any other model set which contains the true system also contains the FSS. We next need to define a norm on the space of systems $\mathcal{S}$, which we take to be the H_∞ norm:

$$||S||_\infty = \sup_{0 \leq \omega \leq 2\pi} |S(\omega)|,$$

where $S(\omega) = \sum_{t=0}^{\infty} h_t^S \exp(-i\omega t)$, which is the z-transform of h^S evaluated at frequency ω.

In this setting, we can view an identification algorithm ϕ as mapping from the observations and assumptions into a *nominal model* $\hat{M}$:

$$\phi(y, x, L, \rho, \varepsilon) = \hat{M} \in \mathcal{S}.$$

Then the *identification error* E of the algorithm or the nominal model is defined as the maximal error over the FSS, given the data:

$$E(\phi) = E(\hat{M}) = \sup_{S \in FSS} ||S - \hat{M}||_\infty.$$

Note that this is worst case with respect to the class of models, but not with respect to the noise realizations. Then we will associate with each algorithm an additive model set $\hat{\mathcal{M}}$ consisting of the nominal model and the unmodeled dynamics:

$$\hat{\mathcal{M}} = \left\{ \hat{M} + \Delta : ||\Delta||_\infty \leq E(\hat{M}) \right\}. \tag{21}$$

An optimal algorithm ϕ^* is one that has the lowest identification error, and it produces as a nominal model the *Chebicheff center* M^c of the FSS:

$$\phi^*(y, x, L, \rho, \varepsilon) = M^c = \arg \inf_{M \in \mathcal{S}} \sup_{S \in FSS} ||S - M||_\infty.$$

Computing the Chebicheff center is a difficult problem with no known general solution. In the next section we describe a "nearly optimal" algorithm that is guaranteed to have identification error at most $\sqrt{2}$ times as large as the theoretical minimum, but the actual extent of suboptimality can be evaluated and is often much lower than the guaranteed bound.

It is worth emphasizing what the estimation procedure delivers. Below we describe an algorithm which maps observed data into a nominal model $\hat{M}$ and a model class $\hat{\mathcal{M}}$ as in (21). This is precisely what is needed for robust control problems, and so SM theory provides a natural formulation to link estimation and control.

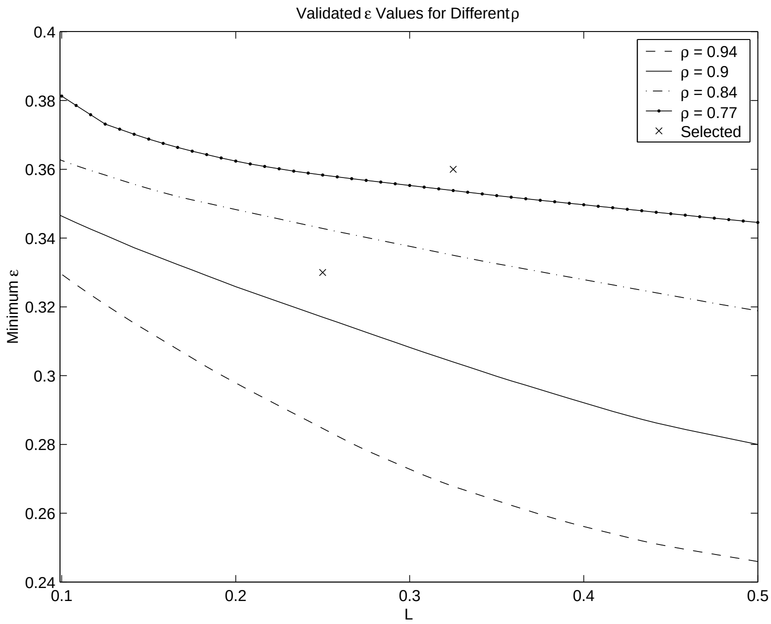

Figure 7: Minimum validated noise bound ε versus the immediate impact L for different values of the decay rate ρ. The selected assumptions are $(L, \rho, \epsilon) = (0.25, 0.9, 0.33)$ and $(0.325, 0.77, 0.36)$.

4.2 The Set Membership Estimation Procedure

With the general SM framework established, we now describe in more detail the estimation procedure, due to Milanese and Taragna (2001) (henceforth MT), that we use. As we noted above, the analysis of the FSS is a key part of the identification algorithm. Recall that the FSS depends on three parameters which must be specified, L which bounds the immediate impact of the input on the output, ρ which gives the decay rate of the impulse responses, and ε which bounds the additive noise. While we cannot "test" these assumptions, we can see if they are violated by the observed data. For a given specification of (L, ρ, ε), if the FSS is non-empty, then the prior assumptions are not violated by the data, and we say that the assumptions are validated.

In particular, there are clear tradeoffs in the specifications of the different parameters. Higher values of L allow for larger responses and higher values of ρ allow for more persistent responses. In either case, more variation in the impulse responses is attributed to the model, and therefore both are consistent with lower bounds on the noise ε. For given values of (L, ρ), MT describe how to find an estimate of the minimum value of ε that is consistent with the data. This gives rise to a surface $\varepsilon(L, \rho)$ which depicts the tradeoffs we just mentioned. Two slices of this surface for our estimation problem are shown in Figures 7 and 8. In Figure 7, we plot $\varepsilon(L, \rho)$ as a function of L for different values of ρ. Here we clearly see that attributing more of the variation in the data to the immediate responses is consistent with lower values of the noise, but that the noise levels decline at a decreasing rate. In Figure 8 we plot the other slice, showing $\varepsilon(L, \rho)$ as a function of ρ for different ε. Here again we see

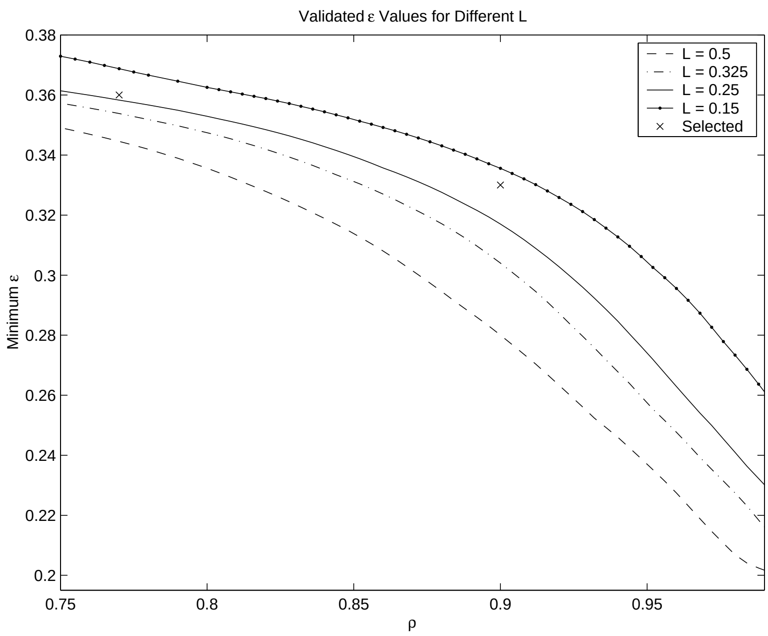

Figure 8: Minimum validated noise bound ε versus the decay rate ρ for different values of the immediate impact L. The selected assumptions are $(L, \rho, \epsilon) = (0.25, 0.9, 0.33)$ and $(0.325, 0.77, 0.36)$.

that attributing more variation in the data to more persistent responses is consistent with less noise, although the levels decline fairly slowly until ρ is very close to 1.

Based on these minimum error surfaces, we arbitrarily choose two cases $(L, \rho) = (0.25, 0.9)$ and $(0.325, 0.77)$. The first set of assumed values falls roughly in the middle of the reasonable range for these parameters, while the second assumes a faster decay of the impulse responses. These specifications are critical, as they directly impact the rest of the results to follow. We see below that the different assumptions give rather different results. Given these choices, the minimum values of the noise bound are calculated to be 0.31 and 0.35, and so we choose the slightly higher values of $\varepsilon = 0.33$ and 0.36. These configurations of parameters are shown on Figures 7 and 8. To gain some further insight into the plausibility of these parameter values, we estimate a number of conventional ARX models using OLS and compare their impulse responses. These are plotted in Figure 9, which shows the impulse responses of all ARX models up to order 11 in the state and order 4 in the input. The bounds on the impulse responses $(\pm L\rho^t)$ are also shown for comparison. Here we see that our assumptions seem quite reasonable, as all of the impulse responses are within the bounds or just barely reach the bounds. When we estimated ARX models of even higher order, we did find larger cycles in the impulse responses which increased in magnitude for the first ten quarters, but these rather implausible results were driven by poorly estimated parameters. Finally, note that the impulse responses of the ARX models all decay to essentially zero well before the end of the 40 quarters shown. As our estimation procedure will rely on finite-length impulse responses, Figure 9 suggests that examining the impact of the output gap on changes in

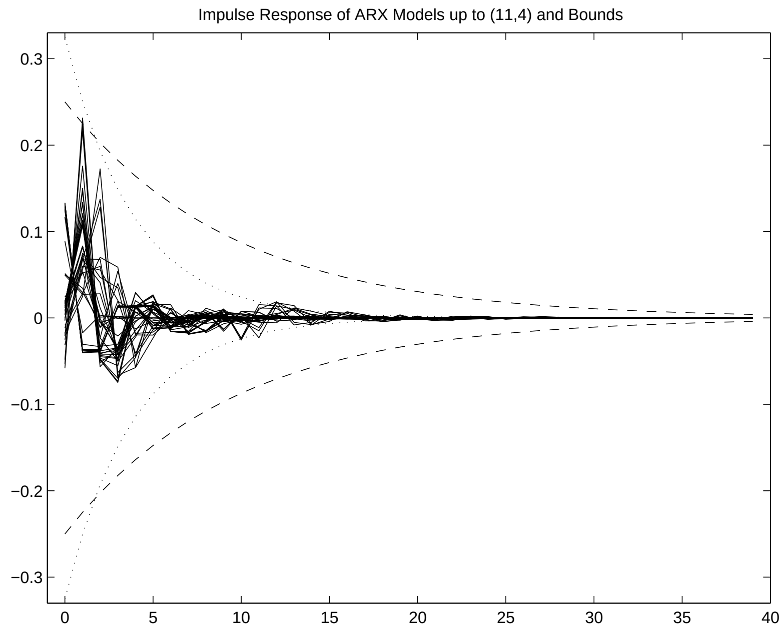

Figure 9: Impulse responses of different ARX models estimated by OLS. The bounds shown correspond to $(L, \rho) = (0.25, 0.9)$ and (0.325,0.77).

inflation over a ten year horizon should be more than sufficient.

With the *a priori* assumptions on the model set and the noise determined, we now turn to identification of the model. Given the specification of the FSS in (20), we can define the value set $V(\omega)$ at a given frequency ω as the set (in the complex plane) of z-transforms of all feasible responses evaluated at ω:

$$V(\omega) = \left\{ (s_1, s_2) \in \mathbb{R}^2 : s = \sum_{t=0}^{\infty} h_t^S \exp(-i\omega t),\ s_1 = \text{Real}(s),\ s_2 = \text{Imag}(s),\ S \in FSS \right\}, \tag{22}$$

where Real$(\cdot)$ and Imag$(\cdot)$ respectively denote the real and imaginary parts of a complex number. The identification procedure is based on approximating the value sets, and then using these approximations to obtain bounds on the identification error of any algorithm. The best algorithm in this case, known as the "nearly optimal" algorithm, minimizes the upper bound on the identification error.

The first step in the approximation is to truncate the impulse responses to a finite length ν, leading to truncated models denoted S^ν. We saw above that in our example the impulse responses are essentially zero after 40 periods, so we choose $\nu = 40$. We then define the following approximations of the FSS:

$$\begin{aligned} \underline{FSS} &= \left\{ S^\nu \in \mathcal{S} : h^{S^\nu} \in H,\ ||x - Yh^{S^\nu}||_\infty \le \varepsilon \right\}, \\ \overline{FSS} &= \left\{ S^\nu \in \mathcal{S} : h^{S^\nu} \in H,\ ||x - Yh^{S^\nu}||_\infty \le \varepsilon + \frac{L\rho^\nu}{1-\rho} \right\}. \end{aligned}$$

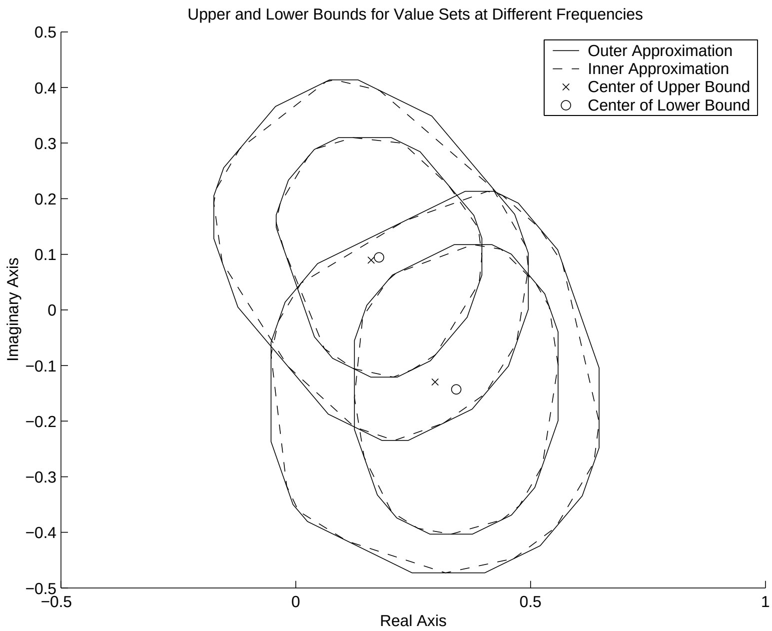

Figure 10: Approximations of the value set $V(\omega)$ for different frequencies. Shown are the inner and outer approximations to the upper and lower bounds, and their corresponding centers.

MT show that by substituting $\underline{FSS}$ and $\overline{FSS}$ instead of the FSS into (22) we obtain lower and upper bounds on $V(\omega)$ respectively. By truncating the impulse responses, the bounds on the value sets are now described by a finite number of inequalities, and thus are polytopes in $\mathbb{R}^2$. MT further describe a convenient numerical method for calculating inner and outer approximations for the polytopes. We refer to the outer approximation to the upper bound (which is clearly a further upper bound) on the value set as $\overline{V}(\omega)$.

Figure 10 shows the inner and outer approximations to the upper and lower bounds on the value sets at two selected frequencies, as well as the Chebicheff centers of the upper and lower bounds. The figure shows the case of $L = 0.25$ and $\rho = 0.9$. First note that the inner and outer approximations are quite close, so the numerical error is small. However the upper and lower bounds differ by a noticeable amount, even though their centers are relatively close. The distance between the bounds is due to the relatively slow decay in the impulse responses. Further, even the lower bounds on the value sets contain a fair amount of area at each frequency, suggesting that there is quite a bit of model uncertainty. This point will recur frequently in the results to follow.

From the approximations to the value set, we can determine a bound on the identification error for any given nominal model $\hat{M}$. In particular, we define the following:

$$\begin{aligned} \overline{E}(\hat{M},\omega) &= ||\overline{V}(\omega) - \hat{M}(\omega)||_2 + \frac{L\rho^{\nu}}{1-\rho}, \\ \overline{E}(\hat{M}) &= \sup_{0\leq\omega\leq 2\pi} \overline{E}(\hat{M},\omega). \end{aligned} \tag{23}$$

Model	Upper Bound on Identification Error		
	$L = 0.25, \rho = 0.9$	$L = 0.325, \rho = 0.77$	
Frequencies	All	All	Business Cycle
Nearly Optimal	1.515	0.632	0.392
Rudebusch-Svensson	1.755	0.762	0.476
OLS ARX(1)	1.700	0.752	0.472
OLS ARX(2)	1.695	0.757	0.474
OLS ARX(4)	1.694	0.762	0.471
Optimal ARX(1)	1.638	0.719	0.468
Optimal ARX(2)	1.625	0.718	0.468
Optimal ARX(4)	1.621	0.718	0.466

Table 4: Upper bound on the identification error for different models under different assumptions. For the ARX models, the number in parentheses refers to the number of lags the changes in inflation.

Then we have the bound $E(\hat{M}) \leq \overline{E}(\hat{M})$. For any nominal model, from (23) we can thus evaluate the identification error and frequency-shaped uncertainty sets.

Similarly, we can choose a nominal model in order to minimize the bound on the identification error. If we restrict ourselves to models of a specific structure, such as ARX models of a given order n, we can optimize over the parameters to determine the optimal model from the given class. We analyze some of these suboptimal models below, which provide models of the form:

$$\mathcal{M}(p^*) = \left\{ M(p^*) + \Delta : \ |\Delta(\omega)| \leq \overline{E}(M(p^*), \omega) \right\} \tag{24}$$

where $p^* = \arg\min \overline{E}(M(p))$. Note that the frequency-weighted uncertainty sets provide a tighter bound on the unmodeled dynamics than the constant bound $||\Delta||_\infty \leq \overline{E}(\hat{M})$ that was assumed in (21) above. While ARX models are tractable and commonly used, there is no need to restrict our attention to this case. We achieve a lower identification error by choosing each of the ν values of the impulse responses h^ν in order to minimize the identification error. MT describe a simple linear programming algorithm which produces such a "nearly optimal" model, whose identification error is guaranteed to be within a factor of $\sqrt{2}$ of the theoretical minimum, but in practice is often substantially lower.

Table 4 reports the upper bound on the identification error $\overline{E}(\hat{M})$ for different nominal models under our two sets of assumptions. In particular, we report the error bounds for the nearly optimal model, the Rudebusch-Svensson model, ARX models estimated by OLS, and optimal ARX models whose parameters have been chosen to minimize the identification error. For the optimal ARX and OLS ARX models, we consider ARX(i, j) models for different orders i in the state, reporting the model with $j \leq i$ that gives the lowest identification error. For the OLS case, we also allowed for the possibility of a delay of one period, as in the RS model. The RS model and OLS ARX models are the same under both sets of assumptions, but their uncertainty is evaluated differently. (The OLS models shown differ with the assumptions since we pick the best model of a given class.) However the nearly optimal models and optimal ARX models differ, as they are tuned to the *a priori*

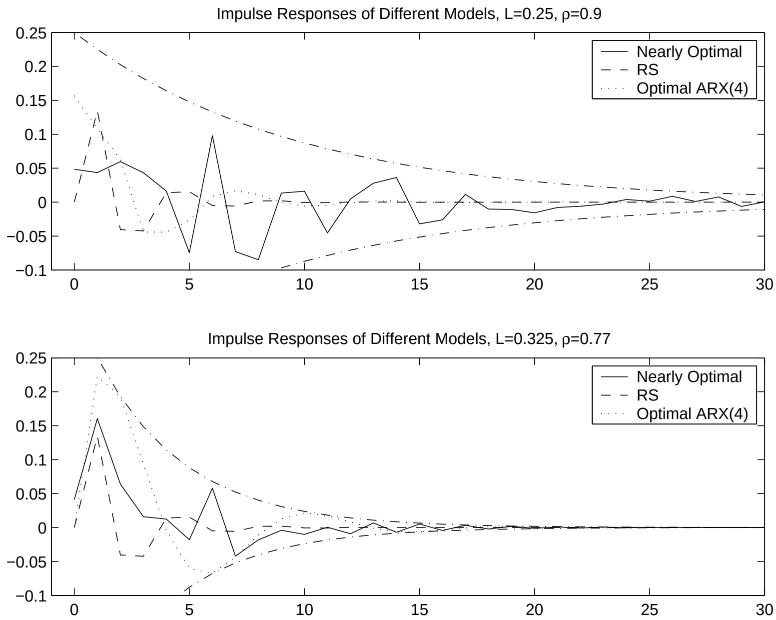

Figure 11: Impulse responses of different models under different assumptions. Shown are the nearly optimal model, the Rudebusch-Svensson model, and the optimal 4th order ARX model, along with the bounds given by $\pm L\rho^t$.

assumptions. Under our second set of assumptions, the table also reports the results when we restrict our estimation procedure to business cycle frequencies (which are defined as above).

Notice that, as expected, the nearly optimal model gives the lowest identification error out of all of the models considered. However the range of variation is not that substantial. Under both sets of assumptions the Rudebusch-Svensson model has the largest identification error, but that its error is only 16-20% larger than the nearly optimal model. Further, the ARX models all performed nearly equally well, with slight improvements possible by optimizing over the parameters, and only very slight decreases in identification error obtained by increasing the model order. Restricting attention to business cycle frequencies also helps in reducing identification error, since (as we show below) the model uncertainty is largest at high frequencies. We defer further discussion of the business cycle frequency results until the end of the section.

The nearly optimal models that we estimate provide noticeable, if modest, reductions in identification error relative to the Rudebusch-Svensson model. However they also imply some substantially different dynamics in the Phillips curve. In Figure 11 we plot the impulse responses of the nearly optimal model, the Rudebusch-Svensson model, and the optimal ARX(4,4) model, along with our *a priori* assumed bounds $\pm L\rho^t$. Recall that these are the impulse responses of changes in inflation to the output gap, not the usual impulse responses to shocks. In the RS model, a one percentage point increase in the output gap leads to an increase in inflation of 0.16 percentage points in the following period, then some slight damped oscillations which decay to nearly zero around quarter 10. Under our first set of

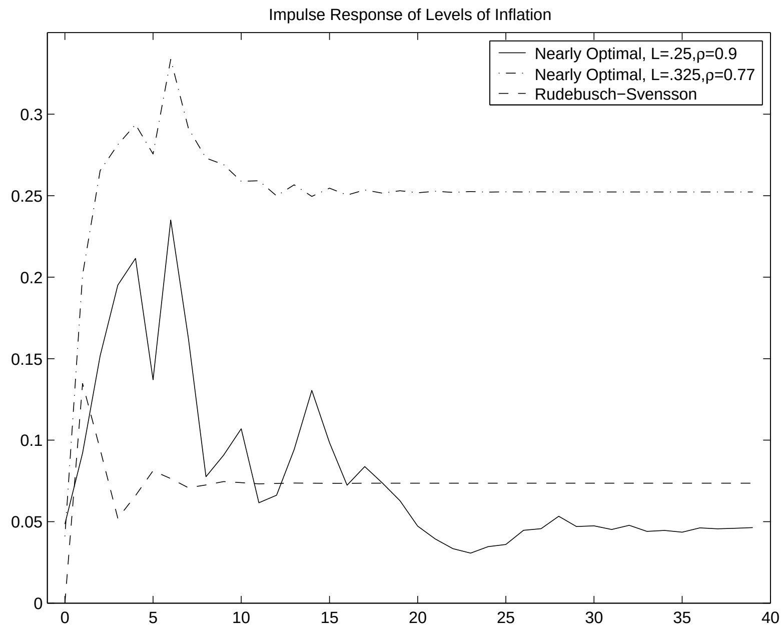

Figure 12: Impulse responses of changes in the output gap to levels of inflation for the nearly optimal and Rudebusch-Svensson models.

assumptions, the nearly optimal model shows that a unit increase in the output gap leads to a simultaneous increase in inflation of 0.05 percentage points, followed by more prolonged and irregular oscillations. The optimal ARX(4,4) model has a much larger simultaneous increase in inflation than the other two, and more prolonged oscillations than the RS model. Under the second set of assumptions, the impulse responses of the models shown have similar qualitative features, but differ greatly in magnitudes. The nearly optimal model is associated with a larger immediate effect and a slower decay than the RS model, while the optimal ARX model has large and persistent oscillations. The implications of the nearly optimal and RS models for *levels* of inflation are shown in Figure 12. There we see that under both sets of assumptions the nearly optimal model associates increases in the output gap with inflation rates that are much higher than the RS model for a sustained period of time. Under our first set of assumptions the permanent increase in inflation is lower than in the RS model, while under the second set the permanent increase is much larger.

In order to gauge how model uncertainty varies at different frequencies, Figure 13 shows the frequency-dependent uncertainty bounds $\overline{E}(\hat{M}, \omega)$. Under both sets of assumptions, we show the bounds for the nearly optimal model, the RS model, and the optimal ARX(4,4) model. Here we see that in all cases, the model uncertainty is smallest at low frequencies, and peaks at very high frequencies. Moreover, the performance of the nearly optimal model is substantially better over the middle range of frequencies, where the ARX models perform worse. Under the first set of assumptions, the optimal ARX model is very close to the RS model over the whole range, with the improvement in the identification error from the optimal ARX model only coming at very high frequencies. Under the second set of assumptions, the

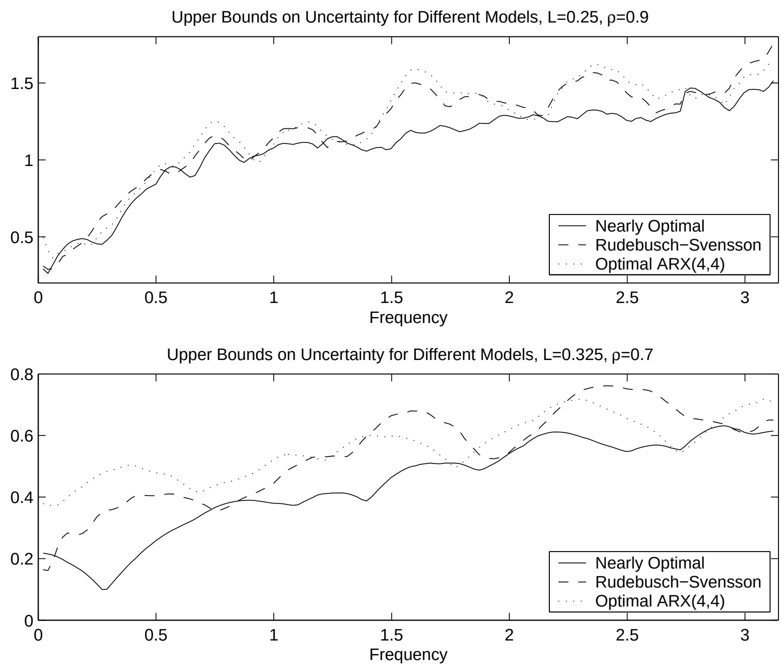

Figure 13: Upper bound on uncertainty as a function of frequency for different models under different assumptions. Shown are the nearly optimal model, the Rudebusch-Svensson model, and the optimal ARX(4,4) model.

RS model outperforms at both relatively low frequencies and very high frequencies.

As in Section 3, we also consider our results when we restrict attention to business cycle frequencies. For this exercise, we only consider the second set of *a priori* assumptions. For the RS model, this involves simply setting the identification error to zero outside the business cycle frequency band. For the nearly optimal and optimized ARX models, we now choose the parameters of the models so that they minimize the identification error over this frequency region. We have already reported the estimation results in Table 4 above, where we see that the identification error falls by roughly 40%. While the model uncertainty is greatest at high frequencies, we saw in Section 3 that the low frequency uncertainty has the largest effect on the optimization problem. We return to this point in the next section.

4.3 Policy Rules

In this section we use the Phillips curve models identified in the previous section in order to derive robust optimal Taylor rules. From our estimation results above we have model sets of the form given in (24), with nonparametric frequency-dependent uncertainty weights. As in Section 3 above, we approximate these weights with rational functions which are then used in the optimization procedures. In order to make the control rules interpretable and consistent with the rest of the paper, we consider simple Taylor-type rules which feedback on the four quarter average inflation rate and the output gap. As our estimation results

Assumptions	Frequencies	Uncertainty Scaling	Coefficient on Inflation	Coefficient on Output Gap	Worst Possible Loss
Nearly Optimal Models					
1	All	0	3.0	2.2	0.57
1	All	10%	3.2	2.6	1.30
2	All	0	2.8	4.2	0.31
2	All	10%	2.8	4.3	0.35
2	All	100%	3.6	4.3	2.87
2	Business	0	2.9	4.0	0.34
2	Business	10%	2.9	4.9	0.36
2	Business	100%	1.6	2.5	0.75
Optimal ARX(4) Models					
1	All	0	2.7	3.3	0.30
1	All	10%	2.6	3.6	0.41
2	All	0	2.6	5.1	0.29
2	All	10%	2.7	5.6	0.34
2	Business	0	3.1	1.9	0.60
2	Business	10%	3.0	2.5	0.71
Rudebusch-Svensson Model					
1,2	Any	0	3.1	2.2	0.55
1	All	10%	3.4	3.1	1.15
2	All	10%	3.4	3.0	0.85
2	Business	10%	3.0	2.7	0.68
Nearly Optimal Model with IS Curve Shocks					
2	All	100%	3.6	5.3	62.19
2	Business	100%	1.7	3.2	17.99

Table 5: The coefficients of the robust optimal Taylor rules and associated worst possible losses for different models under different assumptions. For Assumptions 1, $L = 0.25, \rho = 0.9$. For Assumptions 2, $L = 0.325, \rho = 0.77$.

in the previous section only considered the Phillips curve, we augment the model with the estimated IS equation from the Rudebusch-Svensson model. For most of our results we also zero out the shocks in the IS equation. This provides the simplest method for obtaining an interest rate rule from our Phillips curve estimates.

Once we specify the uncertainty weights and the IS equation, the policy optimization problem is the same as in Section 3 above, but now with different nominal models and different measurements of uncertainty. Then we tune the coefficients of the Taylor-type policy rule in order to minimize the upper bound on the worst case quadratic loss. Unfortunately, under our first set of assumptions the uncertainty that we estimated is simply too large to be of direct use in this optimization. Because we imposed so few *a priori* assumptions and allowed for models with relatively persistent effects, we are left with a very large uncertainty set. Under these assumptions, we could not find any plausible settings of the Taylor rule leading to finite minimax loss for any of the estimated models. However our second set of assumptions implies a much faster decay of the impulse responses, but even in this case only the nearly optimal model (and none of the others) resulted in finite loss.

A simple, if arbitrary, means of using our estimation results is to scale all of the uncertainty weights down by a constant fraction. For illustration, we computed the robust optimal Taylor rules under the assumption that the model uncertainty was only 10% as large as the amount we estimated. For comparison we also computed the results for the different models under no uncertainty, and for the nearly optimal model under the full uncertainty. Our results are shown in Table 5, where we give the coefficients of Taylor-type rules for different nominal models under different assumptions. As in the previous sections, we also report results when we restrict attention to business cycle frequencies. Note that the losses reported in this table are not directly comparable to the results in Section 3 because for the most part we zero out the shocks to the IS curve. To make these comparisons, we also compute results for the nearly optimal model when we reintroduce the IS shocks. Recall that the nearly optimal and optimal ARX models are tuned to the measurement of uncertainty, so the models change depending on our assumptions and the frequency range considered.

From Table 5 we can see some general results. First, the *a priori* assumptions matter greatly. Under the first set of assumptions the uncertainty is quite large. Even allowing for 10% of the measured uncertainty can more than double losses relative to the no uncertainty case (under the nearly optimal and RS models) and makes the policy rules (in all cases) more aggressive in their responses to both the inflation rate and the output gap. Under the second set of assumptions, the uncertainty is more manageable, and allowing for 10% of the measured uncertainty leads to more modest increases in losses and very small increases in aggressiveness. It is notable that only the nearly optimal model, which is designed to minimize model uncertainty can handle the full 100% uncertainty. Allowing for this uncertainty at all frequencies, we see that the losses increase by a factor of more than 9 and the policy rule becomes quite aggressive.

Second, uncertainty which enters at different frequencies has different effects. In most cases, allowing for uncertainty at all frequencies leads to more aggressive policy while restricting attention to business cycle frequencies generally leads to attenuation in policy. This effect is strongest in the nearly optimal model under the second set of assumptions, allowing for the full 100% uncertainty. In this case the policy rule is significantly less aggressive than the no uncertainty case, and has the smallest inflation response coefficient of any of the specifications we consider.

In summary, we obtain our sharpest results by considering model sets with less persistent effects and by focusing on business cycle frequency perturbations. It may be possible to sharpen our results even further by weighting model uncertainty in the estimation stage. Rather than minimizing model uncertainty across frequencies, our estimation procedure would then consider where uncertainty is the most damaging for policy.

5 Conclusion

In this paper we analyzed the effects of uncertainty on monetary policy decisions. We considered four different types of uncertainty: uncertainty about the specification of a reference model, uncertainty about the serial correlation of noise, uncertainty about data quality, and uncertainty about the choice of the reference model. We argued that different specifications of uncertainty may have significantly different implications for monetary policy. Further, uncertainty which enters at different frequencies has substantially different effects. It is therefore necessary to model the uncertainty itself and try to carefully estimate or calibrate the uncertainty model.

We introduced two different ways to systematically approach the formulation of uncertainty relevant for policy making. Our first approach is based on the Model Error Modeling literature. Following this approach, to describe the uncertainty one should first estimate a reference model of the economy. Then take the reference models errors and model them using a relatively unrestricted model. Finally, build the model uncertainty set around the reference model which is consistent with all models for the errors not rejected by formal statistical procedures.

Our second approach exploits recent advances in the Set Membership identification literature. We model uncertainty by an infinite number of linear constraints on the true model's parameters and shocks. In particular, absolute values of shocks are required to be bounded by a given positive number and the impulse responses are required to decay geometrically. The model uncertainty set is represented by a set of the parameters that is not falsified by the data and these linear restrictions. An advantage of the latter approach over the former is that fewer *a priori* restrictions on the true model are introduced. Moreover, it is possible to simultaneously obtain a reference model and uncertainty description using an estimation procedure designed to produce the most compact description of the uncertainty. However, a relative disadvantage of the Set Membership approach is that it can be applied only to one-dimensional input-output models and may result in large model uncertainty sets.

In the numerical part of the paper, we used both approaches to build a description of economic uncertainty with a final goal of designing robust monetary policy rules of Taylor type. We used the minimax criterion of optimality for the policy rules. That is the robust rule was defined as the policy rule that minimizes losses under the worst possible scenario consistent with the uncertainty description. We found that in both cases our description of uncertainty was often too large to produce sensible recommendations for policy makers. Only by imposing a faster rate of decay of the impulse responses in the SM case were we able to tolerate the measured uncertainty. We ranked different types of uncertainty according to their potential destructiveness for the Taylor-type policy rules. The most damaging source of uncertainty for a policy maker is found to be the pure model uncertainty, that is the uncertainty associated with the specification of the reference model. The second worst is the real-time data uncertainty. Finally, the most manageable evil is the pure shock uncertainty.

We also found that although the measured uncertainty is largest at high frequencies, very low frequency perturbations have the most impact on policy. The aggressiveness we found in policy rules is driven almost entirely by worst-case scenarios at very low frequencies. Since our baseline model is essentially a model of short-run fluctuations, we felt that it was extreme to ask it to accommodate very low frequency perturbations. Therefore we recalculated our results by restricting attention to business cycle frequencies. In these cases we found that instead of reacting aggressively, our policy rules were more attenuated than in the absence of uncertainty. Under some specifications, our results were quite close to the policy rules that have been estimated based on actual data.

We started our paper by saying that uncertainty is pervasive in economics. We can now add that in many cases the amount of the uncertainty seems to be unmanageably large. By allowing for low frequency perturbations, policy must respond very aggressively and policy makers face large worst-case losses. However if we restrict attention to business cycle frequencies, the optimally robust policy is less aggressive than the optimal policy under no uncertainty. This finding supports Brainard's conservatism principle.

References

Anderson, E., L. P. Hansen, and T. Sargent (2000). Robustness, detection and the price of risk. Mimeo.

Ball, L. (1999). Policy rules for open economies. In J. Taylor (Ed.), *Monetary Policy Rules*, pp. 127–144. Chicago: University of Chicago Press for the NBER.

Blinder, A. (1997). Distinguished lecture on economics in government: What central bankers could learn from academics - and vice versa. *Journal of Economic Perspectives 11*, 3–19.

Brainard, W. (1967). Uncertainty and the effectiveness of policy. *American Economic Review 57*, 411–425.

Carlson, J. M. and J. Doyle (2002). Complexity and robustness. *Proceedings of the National Academy of Sciences 99*, 2538–2545.

Chamberlain, G. (2000). Econometric applications of maxmin expected utility theory. *Journal of Applied Econometrics 15*, 625–644.

Chib, S. and E. Greenberg (1994). Bayes inference in regression models with ARMA(p,q) errors. *Journal of Econometrics 64*, 183–206.

Chow, G. C. (1975). *Analysis and Control of Dynamic Economic Systems*. New York: John Wiley and Sons.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics 18*, 141–153.

Giordani, P. and P. Söderlind (2002). Solution of macromodels with Hansen-Sargent robust policies. Mimeo.

Hansen, L. P. and T. J. Sargent (2002). *Elements of Robust Control and Filtering for Macroeconomics*. To be decided.

Ljung, L. (1999). Model validation and model error modeling. Control and Communcations Group Technical Report LiTH-ISY-R-2125, Linkoping University.

McCallum, B. T. (1988). Robustness properties of a rule for monetary policy. *Carnegie-Rochester Conference Series on Public Policy 29*, 175–203.

Milanese, M. and M. Taragna (1999). SM indentifcation of model sets for robust control design. In A. Garulli, A. Tesi, and A. Vicino (Eds.), *Robustness in Indentification and Control*, pp. 17–34. New York: Springer-Verlag.

Milanese, M. and M. Taragna (2001). Optimality, approximation and complexity in set membership H_∞ identificaton. Mimeo.

Onatski, A. (2001). Essays on robust policy under model uncertainty. Ph.D. thesis, Harvard University.

Onatski, A. and J. H. Stock (2002, February). Robust monetary policy under model uncertainty in a small model of the US economy. *Macroeconomic Dynamics 6*, 85–110.

Orphanides, A. (2001). Monetary policy rules based on real-time data. *American Economic Review 91*, 964–985.

Paganini, F. (1996). Set constraints in the analysis of uncertain systems. Ph.D. thesis, California Institute of Technology.

Rudebusch, G. and L. E. O. Svensson (1999). Policy rules for inflation targeting. In J. Taylor (Ed.), *Monetary Policy Rules*, pp. 203–246. Chicago: University of Chicago Press for the NBER.

Rudebusch, G. D. (2001, May). Is the Fed too timid? Monetary policy in an uncertain world. *Review of Economics and Statistics 83*, 203–217.

Rust, J. (1997). Using randomization to break the curse of dimensionality. *Econometrica 65*, 487–516.

Rust, J., J. F. Traub, and H. Wozniakowski (2002). Is there a curse of dimensionality for contraction fixed points in the worst case? *Econometrica 70*, 285–329.

Sargent, T. J. (1999). Discussion of "Policy rules for open economies," by Laurence Ball. In J. Taylor (Ed.), *Monetary Policy Rules*, pp. 145–158. Chicago: University of Chicago Press for the NBER.

Sims, C. A. (2001). Pitfalls of a minimax approach to model uncertainty. *American Economic Review 91*, 51–54.

Taylor, J. B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy 39*, 195–214.

Traub, J. F., H. Wasilkowski, and H. Wozniakowski (1988). *Information-Based Complexity.* New York: Academic Press.

Zhou, K., J. Doyle, and K. Glover (1996). *Robust and Optimal Control.* Upper Saddle River, New Jersey: Prentice Hall.