FEDERAL RESERVE BANK OF SAN FRANCISCO

WORKING PAPER SERIES

Robust Control with Commitment: A Modification to Hansen-Sargent

Richard Dennis
Federal Reserve Bank of San Francisco

December 2006

Working Paper 2005-20
http://www.frbsf.org/publications/economics/papers/2005/wp05-20bk.pdf

Robust Control with Commitment: A Modification to Hansen-Sargent*

Richard Dennis†
Federal Reserve Bank of San Francisco

December, 2006

Abstract

This paper examines the Hansen and Sargent (2003) formulation of the robust Stackelberg problem and shows that their method of constructing the approximating equilibrium, which is central to any robust control exercise, is generally invalid. The paper then turns to the Hansen and Sargent (2007) treatment, which, responding to the problems raised in this paper, changes subtly, but importantly, how the robust Stackelberg problem is formulated. This paper proves, first, that their method for obtaining the approximating equilibrium is now equivalent to the one developed in this paper, and, second, that the worst-case specification errors are not subject to a time-inconsistency problem. Analyzing robust monetary policy in two New Keynesian business cycle models, the paper demonstrates that a robust central bank should primarily fear that the supply side of its approximating model is misspecified and that attenuation characterizes robust policymaking. Depending on how the robust Stackelberg problem is formulated, this paper shows that the Hansen-Sargent approximating equilibrium can be dynamically unstable and that robustness can be irrelevant, i.e., that the robust policy can coincide with the rational expectations policy.

Keywords: *Robust control, Robust Stackelberg games, Approximating equilibrium.*

JEL Classification: E52, E62, C61.

*I would like to thank Tom Sargent for discussions on robust control and Eric Swanson and Carl Walsh for comments on an earlier draft of this paper. The views expressed in this paper do not necessarily reflect those of the Federal Reserve Bank of San Francisco or the Federal Reserve System.

†Address for Correspondence: Economic Research, Mail Stop 1130, Federal Reserve Bank of San Francisco, 101 Market St, CA 94105, USA. Email: richard.dennis@sf.frb.org.

1 Introduction

Uncertainty of one form or another is a problem that all decisionmakers must confront, and it is an issue that is clearly important for policy institutions such as the Federal Reserve. In the words of former Federal Reserve chairman Greenspan, "Uncertainty is not just an important feature of the monetary policy landscape; it is the defining characteristic of that landscape" (Greenspan, 2003). Recognizing the impact that uncertainty can have on behavior, Hansen and Sargent and coauthors show how economic agents can use robust control theory to make decisions while guarding against the uncertainty that they fear.[1] Building on and extending rational expectations, robust control allows a concern for model misspecification, or pessimism, to distort how expectations are formed, and thereby alter decisions. Although decisionmakers may possess a "good" model of the economy (their approximating model), fears of misspecification are operationalized by assuming that expectations are formed and decisions are made in the context of a model distorted strategically by specification errors. The distorted model twists the approximating model according to the worst fears of that particular agent.[2] Importantly, although it is common to assume that all agents share the same approximating model, there can be as many distorted models as there are agents in the economy, with each distorted model reflecting the concerns of that particular agent.

It is important to appreciate, however, that although agents may fear that their model is misspecified, specification errors need not actually be present. For this reason, the concept of an approximating equilibrium is central to any robust control exercise. In the approximating equilibrium, although the approximating model is correctly specified, agents that seek robustness apply decision rules that guard against their worst-case specification error. The approximating equilibrium is distinct from the rational expectations equilibrium, in which agents do not fear misspecification, and from an agent's worst-case equilibrium, in which that agent's worst fears materialize. For robust Stackelberg games, Hansen and Sargent (2003) describe a procedure to recover the approximating equilibrium from a pessimistic agent's worst-case equilibrium. However, complicating the analysis of robust Stackelberg games is the fact that the leader's commitment to its policy means that shadow prices, or Lagrange multipliers on

[1]In their manuscript *Robustness*, Hansen and Sargent (2007) provide a sweeping analysis of how robust control can be applied to economic decisionmaking.

[2]Insomuch as in equilibrium they each involve probabilities being distorted, or slanted, toward unfavorable payoffs, there are ties between robust control, risk sensitive control (Whittle, 1990), uncertainty aversion (Epstein and Wang, 1994), disappointment aversion (Gul, 1991), and the latter's generalization (Routledge and Zin, 2003).

implementability conditions, enter the solution.[3]

In this paper, I show that although the method that Hansen and Sargent (2003) use correctly locates the approximating equilibrium for planning problems in which private agents are not forward-looking (Sargent, 1999; Hansen and Sargent, 2001a), it fails for the very class of robust Stackelberg games that they sought to analyze. After demonstrating why their method fails, I show how to recover the approximating equilibrium correctly. Unlike Hansen and Sargent (2003), when recovering the approximating equilibrium, I recognize that the shadow prices that arise in the solution when followers are forward-looking are a component of the optimal policy, they reflect the history dependence of the leader's policy and are intended to induce particular behavior from the followers. Therefore, when constructing the approximating equilibrium, both the leader's decision rule *and* the law of motion for the shadow prices should reflect the leader's fear of misspecification.

To understand the nature of robust monetary policy, I employ two New Keynesian (NK) business cycle models. The first model is the canonical NK model (Clarida, Galí, and Gertler, 1999), which forms the basis of numerous nonrobust optimal policy studies.[4] The second model is the sticky price/sticky wage model developed by Erceg, Henderson, and Levin (2000), which provides the foundation for many modern NK business cycle models. Introducing the specification errors as per Hansen and Sargent (2003), I show that the central bank in the canonical NK model will fear mainly that its approximating model understates the persistence of the shock processes. I also show that, although the Hansen-Sargent approximating equilibrium produces counterintuitive behavior from the central bank and private agents, this unusual behavior will not be brought to light using detection-error methods.[5] For the sticky price/sticky wage model, I show that a pessimistic central bank will fear primarily that its approximating model understates the persistence of the technology shock. And I show that the Hansen-Sargent approximating equilibrium can produce oscillatory behavior or instability, even while both the rational expectations equilibrium and the worst-case equilibrium are stable and non-oscillatory.

After I informed them of these problems,[6] Hansen and Sargent (2007, chapter 16) modified

[3] In this respect, there is a clear parallel between the solution of robust Stackelberg games and nonrobust Stackelberg games, in which shadow prices also serve as state variables (Kydland and Prescott, 1980).

[4] Giordani and Söderlind (2004) also used this model to investigate robust monetary policy. However, discretionary policymaking was their main focus, and their analysis employing commitment was conducted using the Hansen and Sargent (2003) approximating equilibrium.

[5] A detection error occurs when an econometrician observing equilibrium outcomes mistakenly infers which of two models generated the data.

[6] After writing the first version of this paper (April 2005), I sent copies to Lars Hansen and Tom Sargent.

the class of robust Stackelberg games that they analyze; this is an important contribution of this paper. In addition to specification errors residing in the predetermined block of the model, they now allow the Stackelberg leader to fear distortions to private-agent expectations, similar to Woodford (2005). For this alternative formulation of the robust Stackelberg problem, I prove the worst-case distortions are unaffected by the Stackelberg leader's promises about future policy. I also prove that, conditional on the Stackelberg leader's robust decision rule, the worst-case distortions do not affect the worst-case law of motion for the shadow prices. As a consequence, simply setting the worst-case distortions to zero does not alter the law of motion for the shadow prices, and, for this reason, the approximating equilibrium that results is equivalent to the one obtained using my procedure. It follows that, whether interest centers on the robust control problem analyzed in Hansen and Sargent (2003) or on that analyzed in Hansen and Sargent (2007, chapter 16), the method I propose returns the correct approximating equilibrium.

Returning to the two NK business cycle models, allowing the central bank to fear distortions to private-agent expectations, I show that the robust policy in the canonical NK model coincides with the rational expectations policy. In other words, strikingly, the central bank's fear of misspecification has *no* effect on monetary policy, *no* effect on economic outcomes, and *no* effect on welfare. This robustness-irrelevance result arises because the central bank's ability to manipulate expectations allows it to craft a policy that entirely insulates the economy from misspecification. Perhaps fortunately, this robustness-irrelevance result does not carry over to the sticky price/sticky wage model. For that model, I find that the central bank's desire for robustness primarily affects the promises it makes regarding future policy. Across the two models and across the two formulations of the robust control problem, common themes are that the central bank fears distortions to supply shocks and that its desire for robustness does not elicit aggressive policy interventions.

The remainder of the paper is organized as follows. In the following section, I summarize some related literature. In section 3, I introduce the general structure of the approximating model and outline how Hansen and Sargent (2003) formulate the optimization problem confronting a Stackelberg leader who fears misspecification. section 4 shows how control methods developed by Backus and Driffill (1986) can be employed to solve for the worst-case

Subsequently, they acknowledged the error, changed the nature of the distortions that the Stackelberg leader fears, and revised the relevant chapters of their manuscript. As a consequence, I do not believe that the treatment that Hansen and Sargent currently provide in their manuscript, *Robustness*, is in error. Indeed, I show that their method of constructing the approximating equilibrium is now equivalent to mine.

equilibrium.[7] Two methods for constructing the approximating equilibrium are presented and explained in section 5. section 6 clarifies how and why the approximating equilibrium developed in section 5 differs from Hansen and Sargent (2003). section 7 describes and analyzes the Hansen and Sargent (2007) revised formulation of the robust Stackelberg problem. section 8 examines robust monetary policy using two NK business cycle models. section 9 concludes. Appendices contain technical material.

2 Related literature

Prominent applications of robust control have examined how uncertainty affects asset prices. For example, Hansen, Sargent, and Tallarini (1999) study consumption behavior and the price of risk using a model in which a representative consumer, facing a linear production technology and an exogenous endowment process, has a preference for robustness. They show that a preference for robustness operates on asset prices like a change in the discount factor, with expressions for consumption and investment unchanged from the usual permanent income formulas. Hansen, Sargent, and Wang (2002) build on Hansen, Sargent, and Tallarini (1999) by assuming that part of the state vector is unobserved and by introducing robust filtering. Exploiting a separation principle that allows consumers to first estimate the state and then apply robust control as if the state were known, they show that the consumer's filtering problem does not add to the price of risk for a given detection error probability. Cagetti, Hansen, Sargent, and Williams (2002) extend the analysis in Hansen, Sargent, and Wang (2002) to a continuous-time nonlinear stochastic growth model in which technology follows a two-state Markov process.

Where the studies outlined above all employ single-agent robust decision theory, Hansen and Sargent (2003) consider a multi-agent environment, focusing on robust Stackelberg games in which the leader and the follower(s) may each express a preference for robustness separately. To illustrate their approach, Hansen and Sargent (2003) solve a model in which a monopoly leads a competitive fringe. Naturally, robust Stackelberg games can also be applied to multi-agent economies in which a fiscal or a monetary authority seeks to conduct policy while expressing distrust in its model. Taking this path, Giordani and Söderlind (2004) analyze monetary policy in an economy consisting of households, firms, and a central bank. They study robust Stackelberg games (commitment equilibria) and robust Markov-perfect Stackelberg

[7] Backus and Driffill (1986) build on earlier work by Oudiz and Sachs (1985). See also Currie and Levine (1985, 1993).

games (discretionary equilibria) in which the central bank formulates policy while fearing misspecification. In recent work, Hansen and Sargent (2007, chapter 15) consider Markov-perfect equilibria in multi-agent economies, allowing each agent to seek robustness.

3 The benchmark robust Stackelberg game

In this section, I describe a linear-quadratic control problem in which a Stackelberg leader commits to a policy while expressing concern that its model may be misspecified. For simplicity, I focus on an environment in which only one agent — the Stackelberg leader — fears misspecification, and, consistent with Hansen and Sargent (2003), I assume that the specification errors that the Stackelberg leader fears reside in the "predetermined block" of the model.[8] An alternative assumption is to allow the specification errors to distort both the conditional mean and the conditional volatility of the shock processes (Dennis, Leitemo, and Söderström, 2006a). In section 7, I consider the case where the Stackelberg leader also fears that specification errors will distort private-agent expectations, expanding on Hansen and Sargent (2007, chapter 16).

The economic environment is one in which an $n \times 1$ vector of endogenous variables, $\mathbf{z}_t$, consisting of n_1 predetermined variables, $\mathbf{x}_t$, and n_2 ($n_2 = n - n_1$) non-predetermined variables, $\mathbf{y}_t$, evolves over time according to

$$\mathbf{x}_{t+1} = \mathbf{A}_{11}\mathbf{x}_t + \mathbf{A}_{12}\mathbf{y}_t + \mathbf{B}_1\mathbf{u}_t + \mathbf{C}_1\boldsymbol{\varepsilon}_{xt+1}, \tag{1}$$

$$\mathrm{E}_t\mathbf{y}_{t+1} = \mathbf{A}_{21}\mathbf{x}_t + \mathbf{A}_{22}\mathbf{y}_t + \mathbf{B}_2\mathbf{u}_t, \tag{2}$$

where $\mathbf{u}_t$ is a $p \times 1$ vector of policy control variables, $\boldsymbol{\varepsilon}_{xt} \sim iid\,[\mathbf{0}, \mathbf{I}_s]$ is an $s \times 1$ ($s \leq n_1$) vector of white-noise innovations, and E_t is the private sector's mathematical expectations operator conditional upon period t information. Equations (1) and (2) describe the approximating model, the model that the leader and private agents (the followers) believe comes closest to describing the data generating process. The matrices $\mathbf{A}_{11}$, $\mathbf{A}_{12}$, $\mathbf{A}_{21}$, $\mathbf{A}_{22}$, $\mathbf{B}_1$, and $\mathbf{B}_2$ are conformable with $\mathbf{x}_t$, $\mathbf{y}_t$, and $\mathbf{u}_t$ and contain the structural parameters that govern preferences and technology. The matrix $\mathbf{C}_1$ is determined to ensure that $\boldsymbol{\varepsilon}_{xt}$ has the identity matrix as its variance-covariance matrix.

In a rational expectations world, the control problem is for the leader to choose the control

[8] Note that Giordani and Söderlind (2004), Dennis, Leitemo, and Söderström (2006a,b), and numerous other robust control applications model the specification errors this way.

variables $\{\mathbf{u}_t\}_0^\infty$ to minimize

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\mathbf{z}_t^{\prime} \mathbf{W} \mathbf{z}_t + \mathbf{z}_t^{\prime} \mathbf{U} \mathbf{u}_t + \mathbf{u}_t^{\prime} \mathbf{U}^{\prime} \mathbf{z}_t + \mathbf{u}_t^{\prime} \mathbf{R} \mathbf{u}_t \right], \tag{3}$$

where $\beta \in (0,1)$ is the discount factor, subject to equations (1) and (2). The weighting matrices, $\mathbf{W}$ and $\mathbf{R}$, are assumed to be positive semidefinite and positive definite, respectively.

However, the leader is concerned that its approximating model may be misspecified. To accommodate this concern, distortions, or specification errors, $\mathbf{v}_{t+1}$, are introduced and the approximating model is surrounded by a class of models of the form

$$\mathbf{x}_{t+1} = \mathbf{A}_{11} \mathbf{x}_t + \mathbf{A}_{12} \mathbf{y}_t + \mathbf{B}_1 \mathbf{u}_t + \mathbf{C}_1 \left(\mathbf{v}_{t+1} + \boldsymbol{\varepsilon}_{xt+1} \right), \tag{4}$$

$$\mathrm{E}_t \mathbf{y}_{t+1} = \mathbf{A}_{21} \mathbf{x}_t + \mathbf{A}_{22} \mathbf{y}_t + \mathbf{B}_2 \mathbf{u}_t. \tag{5}$$

Equations (4) and (5) describe the "distorted" model, the approximating model distorted by specification errors. Because private agents know that the leader is concerned about misspecification, the distorted model can be written more compactly as

$$\mathbf{z}_{t+1} = \mathbf{A} \mathbf{z}_t + \mathbf{B} \mathbf{u}_t + \mathbf{C} \mathbf{v}_{t+1} + \widetilde{\mathbf{C}} \boldsymbol{\varepsilon}_{t+1}, \tag{6}$$

where $\mathbf{C} \equiv \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{0} \end{bmatrix}$, $\widetilde{\mathbf{C}} \equiv \begin{bmatrix} \mathbf{C}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$, and $\boldsymbol{\varepsilon}_{t+1} \equiv \begin{bmatrix} \boldsymbol{\varepsilon}_{xt+1} \\ \boldsymbol{\varepsilon}_{yt+1} \end{bmatrix}$ in which $\boldsymbol{\varepsilon}_{yt+1} \equiv \mathbf{y}_{t+1} - \mathrm{E}_t \mathbf{y}_{t+1}$ is a Martingale difference sequence; these expectation errors are determined as part of equilibrium.

Assume that the pair ($\mathbf{A}$, $\mathbf{B}$) is stabilizable (Kwakernaak and Sivan, 1972, chapter 6) and that the sequence of specification errors, $\{\mathbf{v}_{t+1}\}_0^\infty$, satisfies the boundedness condition

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^{t+1} \mathbf{v}_{t+1}^{\prime} \mathbf{v}_{t+1} \leq \eta, \tag{7}$$

where $\eta \in [0, \overline{\eta})$. It is the satisfaction of this boundedness condition that defines the sense in which the approximating model, shown in equations (1) and (2), is a "good" one. Note that in the special case in which $\eta = 0$, the nonrobust control problem is restored.

To guard against the misspecifications that it fears, the leader formulates policy subject to the distorted model with the mind-set that the specification errors will be as damaging as possible, a view that is operationalized through the metaphor that $\{\mathbf{v}_{t+1}\}_0^\infty$ is chosen by a fictitious evil agent whose objectives are diametrically opposed to those of the leader. Applying Luenberger's (1969) Lagrange multiplier theorem, Hansen and Sargent (2001b) show that the constraint problem, in which equation (3) is minimized with respect to $\{\mathbf{u}_t\}_0^\infty$ and maximized with respect to $\{\mathbf{v}_{t+1}\}_0^\infty$, subject to equations (6) and (7), can be replaced with an equivalent multiplier problem, in which

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\mathbf{z}_t^{'}\mathbf{W}\mathbf{z}_t + \mathbf{z}_t^{'}\mathbf{U}\mathbf{u}_t + \mathbf{u}_t^{'}\mathbf{U}^{'}\mathbf{z}_t + \mathbf{u}_t^{'}\mathbf{R}\mathbf{u}_t - \beta\theta\mathbf{v}_{t+1}^{'}\mathbf{v}_{t+1}\right], \tag{8}$$

$\theta \in [\underline{\theta}, \infty)$, is minimized with respect to $\{\mathbf{u}_t\}_0^\infty$ and maximized with respect to $\{\mathbf{v}_{t+1}\}_0^\infty$, subject to equation (6). With the following definitions $\widetilde{\mathbf{u}}_t \equiv \begin{bmatrix} \mathbf{u}_t \\ \mathbf{v}_{t+1} \end{bmatrix}$, $\widetilde{\mathbf{B}} \equiv \begin{bmatrix} \mathbf{B} & \mathbf{C} \end{bmatrix}$, $\widetilde{\mathbf{U}} \equiv \begin{bmatrix} \mathbf{U} & \mathbf{0} \end{bmatrix}$, and $\widetilde{\mathbf{R}} \equiv \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & -\beta\theta\mathbf{I} \end{bmatrix}$, the robust problem can be expressed as[9]

$$\min_{\{\mathbf{u}_t\}_0^\infty} \max_{\{\mathbf{v}_{t+1}\}_0^\infty} \mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\mathbf{z}_t^{'}\mathbf{W}\mathbf{z}_t + \mathbf{z}_t^{'}\widetilde{\mathbf{U}}\widetilde{\mathbf{u}}_t + \widetilde{\mathbf{u}}_t^{'}\widetilde{\mathbf{U}}^{'}\mathbf{z}_t + \widetilde{\mathbf{u}}_t^{'}\widetilde{\mathbf{R}}\widetilde{\mathbf{u}}_t\right], \tag{9}$$

subject to the distorted model

$$\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t + \widetilde{\mathbf{B}}\widetilde{\mathbf{u}}_t + \widetilde{\mathbf{C}}\boldsymbol{\varepsilon}_{t+1}. \tag{10}$$

4 The worst-case equilibrium

The problem facing a Stackelberg leader that must conduct policy while fearing that its model is misspecified is described by equations (9) and (10). The solution to this problem returns the leader's worst-case equilibrium, the decision rules, and the law of motion for the state variables that govern the economy's behavior according to the leader's worst-case fears. Because the control problem is linear-quadratic, the leader's worst-case equilibrium can be obtained by applying the Backus and Driffill (1986) solution method, developed originally to solve problems with rational expectations, which formulates the optimization problem as a dynamic program.[10]

Since the control problem is linear-quadratic, the value function takes the form $V(\mathbf{z}_t) = \mathbf{z}_t^{'}\mathbf{V}\mathbf{z}_t + d$ and the dynamic program can be written as

$$\mathbf{z}_t^{'}\mathbf{V}\mathbf{z}_t + d \equiv \min_{\mathbf{u}_t}\max_{\mathbf{v}_{t+1}} \left[\mathbf{z}_t^{'}\mathbf{W}\mathbf{z}_t + \mathbf{z}_t^{'}\widetilde{\mathbf{U}}\widetilde{\mathbf{u}}_t + \widetilde{\mathbf{u}}_t^{'}\widetilde{\mathbf{U}}^{'}\mathbf{z}_t + \widetilde{\mathbf{u}}_t^{'}\widetilde{\mathbf{R}}\widetilde{\mathbf{u}}_t + \beta\mathrm{E}_t\left(\mathbf{z}_{t+1}^{'}\mathbf{V}\mathbf{z}_{t+1} + d\right)\right], \tag{11}$$

subject to

$$\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t + \widetilde{\mathbf{B}}\widetilde{\mathbf{u}}_t + \widetilde{\mathbf{C}}\boldsymbol{\varepsilon}_{t+1}. \tag{12}$$

[9] In moving from the constraint problem to the multiplier problem, the assumptions that the pair ($\mathbf{A}$, $\mathbf{B}$) is stabilizable and the sequence $\{\mathbf{v}_{t+1}\}_0^\infty$ is bounded are replaced with the assumptions that the pair ($\mathbf{A}$, $\widetilde{\mathbf{B}}$) is stabilizable and that $\theta\mathbf{I} - \widetilde{\mathbf{C}}^{'}\mathbf{V}\widetilde{\mathbf{C}}$ is positive definite, where $\mathbf{V}$ is defined implicitly by equation (11). The second of these two assumptions determines the lower bound, $\underline{\theta}$, that θ must exceed.

[10] Alternatively, this robust control problem can be solved using a Lagrangian, or by applying the recursive saddle-point theorem (Marcet and Marimon, 1999).

To keep the paper self-contained, I describe the Backus-Driffill solution method in Appendix A, where I show that in the worst-case equilibrium the law of motion for the state variables is

$$\begin{bmatrix} \mathbf{x}_{t+1} \\ \mathbf{p}_{t+1} \end{bmatrix} = \mathbf{T}\left(\mathbf{A} - \mathbf{B}\mathbf{F}_u - \mathbf{C}\mathbf{F}_v\right)\mathbf{T}^{-1} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{p}_t \end{bmatrix} + \mathbf{C}\varepsilon_{xt+1}, \tag{13}$$

$\mathbf{p}_0 = \mathbf{0}$, where $\mathbf{p}_t$ is an $n_2 \times 1$ vector of shadow prices of the non-predetermined variables, $\mathbf{y}_t$. These shadow prices, $\mathbf{p}_t$, are predetermined variables that enter the equilibrium as state variables to give private agents the necessary incentives to form expectations and to make decisions that conform with the unique stable rational expectations equilibrium that is consistent with the leader's chosen policy. An alternative way of thinking about these shadow prices is that they encode the history dependence of the leader's policy, a history dependence arising from the leader's commitment to its policy. The matrix $\mathbf{T}$ provides the mapping between the state variables, $\mathbf{x}_t$ and $\mathbf{p}_t$, and the endogenous variables, $\mathbf{z}_t$, and is given by $\mathbf{T} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}$, where $\mathbf{V}_{21}$ and $\mathbf{V}_{22}$ are submatrices of $\mathbf{V}$, conformable with $\mathbf{x}_t$ and $\mathbf{y}_t$ (see Appendix A).

Together with the law of motion for the state variables, the solution for the non-predetermined variables is

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{V}_{22}^{-1}\mathbf{V}_{21} & \mathbf{V}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{p}_t \end{bmatrix}, \tag{14}$$

while the robust decision rule(s) for the leader and the worst-case distortions are given by

$$\begin{bmatrix} \mathbf{u}_t \\ \mathbf{v}_{t+1} \end{bmatrix} = -\begin{bmatrix} \mathbf{F}_u \\ \mathbf{F}_v \end{bmatrix} \mathbf{T}^{-1} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{p}_t \end{bmatrix}. \tag{15}$$

The leader's worst-case fear is that the approximating model is distorted by the specification errors contained within equation (15).

5 The approximating equilibrium

The approximating equilibrium describes the economy's behavior under robust decisionmaking but in the absence of misspecification.[11] Although the worst-case equilibrium and the approximating equilibrium differ, because the distorted model is a device the leader employs to achieve robustness, the approximating equilibrium and the leader's worst-case equilibrium

[11] Here the Stackelberg leader is the only agent seeking robustness and, as a consequence, there is a single set of worst-case distortions and a single worst-case equilibrium. More generally, with multiple decisionmakers there is a set of worst-case distortions and a worst-case equilibrium for each agent seeking robustness. Nevertheless, even when multiple agents seek robustness, provided all agents share the same approximating model, there is a single shared approximating equilibrium.

are closely connected. To construct the approximating equilibrium, I begin by taking the worst-case equilibrium, equations (13) – (15) and writing it out in terms of states as

$$\mathbf{x}_{t+1} = \mathbf{M}_{11}\mathbf{x}_t + \mathbf{M}_{12}\mathbf{p}_t + \mathbf{C}_1\varepsilon_{xt+1}, \tag{16}$$

$$\mathbf{p}_{t+1} = \mathbf{M}_{21}\mathbf{x}_t + \mathbf{M}_{22}\mathbf{p}_t, \tag{17}$$

$$\mathbf{y}_t = \mathbf{H}_{21}\mathbf{x}_t + \mathbf{H}_{22}\mathbf{p}_t, \tag{18}$$

$$\mathbf{u}_t = \widetilde{\mathbf{F}}_{u1}\mathbf{x}_t + \widetilde{\mathbf{F}}_{u2}\mathbf{p}_t, \tag{19}$$

$$\mathbf{v}_{t+1} = \widetilde{\mathbf{F}}_{v1}\mathbf{x}_t + \widetilde{\mathbf{F}}_{v2}\mathbf{p}_t, \tag{20}$$

where $\mathbf{H}_{21} \equiv \mathbf{V}_{22}^{-1}\mathbf{V}_{21}$, $\mathbf{H}_{22} \equiv \mathbf{V}_{22}^{-1}$, $\begin{bmatrix} \widetilde{\mathbf{F}}_{u1} & \widetilde{\mathbf{F}}_{u2} \end{bmatrix} \equiv -\mathbf{F}_u\mathbf{T}^{-1}$, $\begin{bmatrix} \widetilde{\mathbf{F}}_{v1} & \widetilde{\mathbf{F}}_{v2} \end{bmatrix} \equiv -\mathbf{F}_v\mathbf{T}^{-1}$, and $\mathbf{M} = \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix} \equiv \mathbf{T}\left(\mathbf{A} - \mathbf{B}\mathbf{F}_u - \mathbf{C}\mathbf{F}_v\right)\mathbf{T}^{-1}$.

With the worst-case equilibrium given by equations (16) – (20), identical representations of the approximating equilibrium can be derived by working with either the approximating model or the distorted model. Building up from the approximating model, the first step is to set aside the worst-case distortions (the evil agent's decision rules) by setting $\widetilde{\mathbf{F}}_{v1} = \mathbf{0}$ and $\widetilde{\mathbf{F}}_{v2} = \mathbf{0}$ in equation (20). The second step is to substitute equations (18) and (19) into the approximating model, equation (1). With these substitutions, in the approximating equilibrium the predetermined variables, $\mathbf{x}_t$, evolve over time according to

$$\begin{aligned}
\mathbf{x}_{t+1} &= \mathbf{A}_{11}\mathbf{x}_t + \mathbf{A}_{12}\mathbf{y}_t + \mathbf{B}_1\mathbf{u}_t + \mathbf{C}_1\varepsilon_{xt+1}, \\
&= \mathbf{A}_{11}\mathbf{x}_t + \mathbf{A}_{12}\left(\mathbf{H}_{21}\mathbf{x}_t + \mathbf{H}_{22}\mathbf{p}_t\right) + \mathbf{B}_1\left(\widetilde{\mathbf{F}}_{u1}\mathbf{x}_t + \widetilde{\mathbf{F}}_{u2}\mathbf{p}_t\right) + \mathbf{C}_1\varepsilon_{xt+1}, \\
&= \left(\mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{H}_{21} + \mathbf{B}_1\widetilde{\mathbf{F}}_{u1}\right)\mathbf{x}_t + \left(\mathbf{A}_{12}\mathbf{H}_{22} + \mathbf{B}_1\widetilde{\mathbf{F}}_{u2}\right)\mathbf{p}_t + \mathbf{C}_1\varepsilon_{xt+1},
\end{aligned} \tag{21}$$

with the remainder of the approximating equilibrium given by equations (17) – (19).

Alternatively, simplifying down from the distorted model, substituting equations (18) – (20) into the distorted model, equation (4), yields

$$\begin{aligned}
\mathbf{x}_{t+1} &= \mathbf{A}_{11}\mathbf{x}_t + \mathbf{A}_{12}\mathbf{y}_t + \mathbf{B}_1\mathbf{u}_t + \mathbf{C}_1\mathbf{v}_{t+1} + \mathbf{C}_1\varepsilon_{xt+1}, \\
&= \mathbf{A}_{11}\mathbf{x}_t + \mathbf{A}_{12}\left(\mathbf{H}_{21}\mathbf{x}_t + \mathbf{H}_{22}\mathbf{p}_t\right) + \mathbf{B}_1\left(\widetilde{\mathbf{F}}_{u1}\mathbf{x}_t + \widetilde{\mathbf{F}}_{u2}\mathbf{p}_t\right) + \mathbf{C}_1\left(\widetilde{\mathbf{F}}_{v1}\mathbf{x}_t + \widetilde{\mathbf{F}}_{v2}\mathbf{p}_t\right) + \mathbf{C}_1\varepsilon_{xt+1}, \\
&= \left(\mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{H}_{21} + \mathbf{B}_1\widetilde{\mathbf{F}}_{u1} + \mathbf{C}_1\widetilde{\mathbf{F}}_{v1}\right)\mathbf{x}_t + \left(\mathbf{A}_{12}\mathbf{H}_{22} + \mathbf{B}_1\widetilde{\mathbf{F}}_{u2} + \mathbf{C}_1\widetilde{\mathbf{F}}_{v2}\right)\mathbf{p}_t + \mathbf{C}_1\varepsilon_{xt+1}.
\end{aligned} \tag{22}$$

Then, zeroing-out the distortions by setting $\widetilde{\mathbf{F}}_{v1} = \mathbf{0}$ and $\widetilde{\mathbf{F}}_{v2} = \mathbf{0}$ in equation (22), produces

$$\mathbf{x}_{t+1} = \left(\mathbf{A}_{11} + \mathbf{A}_{12}\mathbf{H}_{21} + \mathbf{B}_1\widetilde{\mathbf{F}}_{u1}\right)\mathbf{x}_t + \left(\mathbf{A}_{12}\mathbf{H}_{22} + \mathbf{B}_1\widetilde{\mathbf{F}}_{u2}\right)\mathbf{p}_t + \mathbf{C}_1\varepsilon_{xt+1}, \tag{23}$$

with the remainder of the system behaving according to equations (17) – (19). Equation (23) is, of course, identical to equation (21).

Importantly, in the approximating equilibrium, the non-predetermined variables, $\mathbf{y}_t$, the leader's decision variables, $\mathbf{u}_t$, and the shadow prices of the non-predetermined variables, $\mathbf{p}_t$, all respond to the state variables as they do in the worst-case equilibrium. The approximating equilibrium and the worst-case equilibrium differ solely because the absence of misspecification alters the law of motion for the predetermined variables, $\mathbf{x}_t$.

6 The Hansen-Sargent approximating equilibrium

In contrast to the approach above, Hansen and Sargent (2003) derive the approximating equilibrium as follows. Given the leader's worst-case equilibrium, Hansen and Sargent (2003, p. 592) set $\mathbf{F}_v = \mathbf{0}$ in equations (13) and (15) to obtain what I will term the "Hansen-Sargent approximating equilibrium." With this simplification, the law of motion for the state variables becomes

$$\begin{bmatrix} \mathbf{x}_{t+1} \\ \mathbf{p}_{t+1} \end{bmatrix} = \mathbf{T}\left(\mathbf{A} - \mathbf{B}\mathbf{F}_u\right)\mathbf{T}^{-1} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{p}_t \end{bmatrix} + \mathbf{C}\boldsymbol{\varepsilon}_{xt+1}, \tag{24}$$

(see Hansen and Sargent, 2003, equation (3.21a)), while the non-predetermined variables and the leader's decision rule(s) continue to be given by equations (18) and (19), respectively. According to the Hansen-Sargent approximating equilibrium, the decision rules for private agents and for the Stackelberg leader are formulated to guard against the leader's worst-case misspecification, but the law of motion for the state variables — the predetermined variables *and* the shadow prices of the non-predetermined variables — is modified because the specification errors that the leader fears are absent.

On a mechanical level, the difference between the approximating equilibrium presented in section 5 and the Hansen-Sargent approximating equilibrium described by equation (24) is that, whereas I zero out the specification errors by setting $\widetilde{\mathbf{F}}_{v1} = \mathbf{0}$ and $\widetilde{\mathbf{F}}_{v2} = \mathbf{0}$ in equations (20) and (22), Hansen and Sargent (2003) zero out the specification errors by setting $\mathbf{F}_v = \mathbf{0}$ in equations (13) and (15). But whereas $\widetilde{\mathbf{F}}_{v1}$ and $\widetilde{\mathbf{F}}_{v2}$ represent coefficients on the state variables (the predetermined variables and the shadow prices) in the evil agent's decision rule, $\mathbf{F}_v$ describes how the evil agent's decision variables relate to the predetermined variables, $\mathbf{x}_t$, and the non-predetermined variables, $\mathbf{y}_t$. It is apparent from equation (13) that setting $\mathbf{F}_v = \mathbf{0}$ can change the behavior of *both* $\mathbf{x}_t$ *and* $\mathbf{p}_t$, allowing the absence of misspecification to alter the law of motion for the shadow prices even while the decision rules for private agents and the leader are unaltered.

Unfortunately, the method that Hansen and Sargent (2003) propose will return an incorrect solution when setting $\mathbf{F}_v = \mathbf{0}$ changes the law of motion for the shadow prices. The method is generally inappropriate because it treats the shadow prices as if they are physical states, rather than recognizing that they enter the equilibrium to induce private agents to behave in a manner consistent with the equilibrium desired by the Stackelberg leader. Since the Stackelberg leader fears misspecification, and this fear does not depend on whether the approximating model is actually misspecified, in the approximating equilibrium the shadow prices, which encode the history dependence of the leader's robust policy, should respond to the state variables as they do in the worst-case equilibrium.

7 An alternative robust Stackelberg game

Responding to the issues raised in this paper, Hansen and Sargent (2007, chapter 16) change how they formulate the robust Stackelberg problem; the change is subtle, but important. In this section, I describe and interpret this alternative formulation. Then I prove, first, that, conditional on the Stackelberg leader's worst-case decision rule, the worst-case law of motion for the shadow prices of the non-predetermined variables does not depend on the worst-case distortions and, second, that the Stackelberg leader cannot influence the worst-case distortions through its promises about future policy. The former result implies that Hansen and Sargent's (2003) approach to constructing the approximating equilibrium would have been correct had it been applied to the Hansen and Sargent (2007) formulation of the problem. The latter result implies that the trade-offs confronting the metaphorical evil agent are very simple, unencumbered by promises about future policy. As a consequence, the solution is invariant to whether the worst-case distortions are determined once and for all or period by period.

Recall that the approximating model is

$$\mathbf{x}_{t+1} = \mathbf{A}_{11}\mathbf{x}_t + \mathbf{A}_{12}\mathbf{y}_t + \mathbf{B}_1\mathbf{u}_t + \mathbf{C}_1\boldsymbol{\varepsilon}_{xt+1}, \tag{25}$$

$$\mathrm{E}_t\mathbf{y}_{t+1} = \mathbf{A}_{21}\mathbf{x}_t + \mathbf{A}_{22}\mathbf{y}_t + \mathbf{B}_2\mathbf{u}_t, \tag{26}$$

where the expectation operator in equation (26) reflects expectations formed by private agents. Introducing the expectational error, $\boldsymbol{\varepsilon}_{yt+1} \equiv \mathbf{y}_{t+1} - \mathrm{E}_t\mathbf{y}_{t+1}$, the approximating model can be written in terms of realized values

$$\mathbf{x}_{t+1} = \mathbf{A}_{11}\mathbf{x}_t + \mathbf{A}_{12}\mathbf{y}_t + \mathbf{B}_1\mathbf{u}_t + \mathbf{C}_1\varepsilon_{xt+1}, \tag{27}$$

$$\mathbf{y}_{t+1} = \mathbf{A}_{21}\mathbf{x}_t + \mathbf{A}_{22}\mathbf{y}_t + \mathbf{B}_2\mathbf{u}_t + \varepsilon_{yt+1}. \tag{28}$$

Recognizing that in equilibrium the expectational errors will be a linear function of the shocks, $\varepsilon_{yt} = \mathbf{C}_2\varepsilon_{xt}$, equations (27) and (28) can be written more compactly as

$$\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t + \mathbf{B}\mathbf{u}_t + \overline{\mathbf{C}}\varepsilon_{xt+1}, \tag{29}$$

where $\overline{\mathbf{C}} \equiv \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{bmatrix}$. With $\mathbf{C}_2$ yet to be determined, Hansen and Sargent (2007, chapter 16) introduce specification errors to equation (29) and write the distorted model as

$$\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t + \mathbf{B}\mathbf{u}_t + \overline{\mathbf{C}}\mathbf{v}_{t+1} + \overline{\mathbf{C}}\varepsilon_{xt+1}, \tag{30}$$

where the specification errors are, of course, constrained to satisfy equation (7).

Conditional on $\mathbf{C}_2$, the dynamic program for this robust control problem is

$$\mathbf{z}_t'\mathbf{V}\mathbf{z}_t + d \equiv \min_{\mathbf{u}_t}\max_{\mathbf{v}_{t+1}} \left[\mathbf{z}_t'\mathbf{W}\mathbf{z}_t + \mathbf{z}_t'\widetilde{\mathbf{U}}\widetilde{\mathbf{u}}_t + \widetilde{\mathbf{u}}_t'\widetilde{\mathbf{U}}'\mathbf{z}_t + \widetilde{\mathbf{u}}_t'\widetilde{\mathbf{R}}\widetilde{\mathbf{u}}_t + \beta \mathrm{E}_t\left(\mathbf{z}_{t+1}'\mathbf{V}\mathbf{z}_{t+1} + d\right)\right], \tag{31}$$

subject to

$$\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t + \mathbf{B}\mathbf{u}_t + \overline{\mathbf{C}}\mathbf{v}_{t+1} + \overline{\mathbf{C}}\varepsilon_{xt+1}. \tag{32}$$

To obtain the worst-case equilibrium, the approach is to conjecture a solution for $\mathbf{C}_2$ and use the method described in Appendix A to solve the dynamic programming problem. The conjecture for $\mathbf{C}_2$ is then updated according to $\mathbf{C}_2 \leftarrow -\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{C}_1$, see equation (A11), and the procedure is iterated until convergence.[12]

How does this formulation of the robust Stackelberg problem differ from the one described in section 3? The difference lies in how the specification errors distort the economy. In the benchmark formulation described in section 3, the Stackelberg leader fears specification errors that reside in the predetermined block of the model (equation (1)). In contrast, in this section the Stackelberg leader also fears distortions to private-agent expectations, i.e., the Stackelberg leader fears that the expectations that private agents form will be slanted in a direction that makes its stabilization problem more difficult. Put differently, here the Stackelberg leader fears that the followers will use the distorted model to form expectations, and formulates policy accordingly, whereas in section 3 the Stackelberg leader understands that the followers

[12] Although it is not obvious that this procedure will necessarily converge, I did not encounter any difficulties with it in this paper.

will use the approximating model to form expectations, and formulates policy accordingly. Of course, because the solution to this problem involves repeated application of the same method used in section 4, the worst-case equilibrium still takes the form of equations (13) – (15), with the worst-case specification errors distorting the conditional mean of the shock processes.

Lemma 1: Conditional on the Stackelberg leader's robust decision rule, the worst-case law of motion for the shadow prices of the non-predetermined variables does not depend on the worst-case distortions.

Proof: The shadow prices of the non-predetermined variables are given by

$$\mathbf{p}_{t+1} = \begin{bmatrix} \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \mathbf{z}_{t+1}, \tag{33}$$

$$= \begin{bmatrix} \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \left(\mathbf{A}\mathbf{z}_t + \mathbf{B}\mathbf{u}_t + \overline{\mathbf{C}}\mathbf{v}_{t+1} + \overline{\mathbf{C}}\boldsymbol{\varepsilon}_{xt+1} \right). \tag{34}$$

However, the coefficient on the specification errors in this law of motion is given by

$$\begin{bmatrix} \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \overline{\mathbf{C}} = \begin{bmatrix} \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{C}_1 \\ -\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{C}_1 \end{bmatrix} = \mathbf{0}. \ \blacksquare \tag{35}$$

An immediate implication of Lemma 1 is that the law of motion for the shadow prices of the non-predetermined variables is unaffected by whether the approximating model is misspecified or not, the essential property emphasized in this paper. It follows that the procedures described in sections 5 and 6 will deliver the same approximating equilibrium.

Lemma 2: The worst-case distortions do not depend on the non-predetermined variables, precluding the Stackelberg leader from influencing the worst-case distortions through promises about future policy.

Proof: The easiest way to see this important feature of the solution is to reformulate the robust control problem in terms of a Lagrangian and to exploit the relationship between the two optimization problems.

The Lagrangian for the robust Stackelberg game is

$$\begin{aligned} L &= \mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t [\mathbf{z}_t^{\prime}\mathbf{W}\mathbf{z}_t + \mathbf{z}_t^{\prime}\mathbf{U}\mathbf{u}_t + \mathbf{u}_t^{\prime}\mathbf{U}^{\prime}\mathbf{z}_t + \mathbf{u}_t^{\prime}\mathbf{R}\mathbf{u}_t - \beta\theta\mathbf{v}_{t+1}^{\prime}\mathbf{v}_{t+1} \\ &\quad + \boldsymbol{\lambda}_{t+1}^{\prime} \left(\mathbf{A}\mathbf{z}_t + \mathbf{B}\mathbf{u}_t + \overline{\mathbf{C}}\mathbf{v}_{t+1} + \overline{\mathbf{C}}\boldsymbol{\varepsilon}_{xt+1} - \mathbf{z}_{t+1} \right)], \end{aligned} \tag{36}$$

where $\{\mathbf{u}_t\}_0^{\infty}$ is chosen to minimize equation (36) and $\{\mathbf{v}_{t+1}\}_0^{\infty}$ is chosen to maximize equation (36) and where $\boldsymbol{\lambda}_{t+1} \equiv \begin{bmatrix} \boldsymbol{\lambda}_{t+1}^{\mathbf{x}^{\prime}} & \boldsymbol{\lambda}_{t+1}^{\mathbf{y}^{\prime}} \end{bmatrix}^{\prime}$ is a vector of Lagrange multipliers. Focusing on the maximization problem, the first-order condition associated with $\mathbf{v}_{t+1}$ is[13]

$$\frac{\partial L}{\partial \mathbf{v}_{t+1}} = -2\beta\theta\mathbf{v}_{t+1} + \overline{\mathbf{C}}^{\prime}\mathrm{E}_t\boldsymbol{\lambda}_{t+1} = \mathbf{0}. \tag{37}$$

[13] Since the constraints are linear and $\theta > 0$, this first-order condition is associated with a maximum.

As is well known, the Lagrange multipliers in equation (36) are equivalent to the shadow prices on $\mathbf{z}_t$ in the dynamic program, equations (31) and (32). Therefore,[14]

$$\begin{bmatrix} \boldsymbol{\lambda}^{\mathbf{x}}_{t+1} \\ \boldsymbol{\lambda}^{\mathbf{y}}_{t+1} \end{bmatrix} = 2 \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{t+1} \\ \mathbf{y}_{t+1} \end{bmatrix}. \tag{38}$$

Substituting equation (38) into equation (37), this first-order condition can be written as

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{v}_{t+1}} &= -\beta\theta \mathbf{v}_{t+1} + \overline{\mathbf{C}}' \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \mathrm{E}_t \begin{bmatrix} \mathbf{x}_{t+1} \\ \mathbf{y}_{t+1} \end{bmatrix} = \mathbf{0}, & (39) \\ &= -\beta\theta \mathbf{v}_{t+1} + \left(\mathbf{C}_1' \mathbf{V}_{11} + \mathbf{C}_2' \mathbf{V}_{21}\right) \mathrm{E}_t \mathbf{x}_{t+1} + \left(\mathbf{C}_1' \mathbf{V}_{12} + \mathbf{C}_2' \mathbf{V}_{22}\right) \mathrm{E}_t \mathbf{y}_{t+1} = \mathbf{0}. & (40) \end{aligned}$$

The next step is to recognize that $\mathbf{C}_2 = -\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{C}_1$ in equilibrium. Therefore, equation (40) becomes

$$\frac{\partial L}{\partial \mathbf{v}_{t+1}} = -\beta\theta \mathbf{v}_{t+1} + \mathbf{C}_1' \left(\mathbf{V}_{11} - \mathbf{V}_{21}' \mathbf{V}_{22}^{-1} \mathbf{V}_{21}\right) \mathrm{E}_t \mathbf{x}_{t+1} + \mathbf{C}_1' \left(\mathbf{V}_{12} - \mathbf{V}_{21}'\right) \mathrm{E}_t \mathbf{y}_{t+1} = \mathbf{0}, \tag{41}$$

which, because $\mathbf{V}$ is symmetric (see equation (A7)), collapses to

$$\frac{\partial L}{\partial \mathbf{v}_{t+1}} = -\beta\theta \mathbf{v}_{t+1} + \mathbf{C}_1' \left(\mathbf{V}_{11} - \mathbf{V}_{21}' \mathbf{V}_{22}^{-1} \mathbf{V}_{21}\right) \mathrm{E}_t \mathbf{x}_{t+1} = \mathbf{0}. \tag{42}$$

Making $\mathbf{v}_{t+1}$ the subject, equation (42) implies

$$\mathbf{v}_{t+1} = \frac{1}{\beta\theta} \mathbf{C}_1' \left(\mathbf{V}_{11} - \mathbf{V}_{21}' \mathbf{V}_{22}^{-1} \mathbf{V}_{21}\right) \mathrm{E}_t \mathbf{x}_{t+1}, \tag{43}$$

establishing that the worst-case distortions do not depend on $\mathrm{E}_t \mathbf{y}_{t+1}$ and therefore do not depend on $\mathrm{E}_t \boldsymbol{\lambda}^{\mathbf{y}}_{t+1}$ ($\mathrm{E}_t \mathbf{p}_{t+1}$).[15] ■

Lemma 2 has several important implications. First, equation (43) establishes that when determining how to distort the model the fictitious evil agent need only consider the effect its distortions have on the natural state variables. Second, equation (42) shows that when distorting the model the fictitious evil agent equates the marginal cost of a change in its

[14] Note that the "2" in equation (38) is simply a scale factor that is sometimes accommodated by placing a "2" in front of the Lagrange multipliers in equation (36) (which makes the Lagrange multipliers equal to half the shadow prices). Naturally, both approaches yield the same answer.

[15] If this last point is unclear, simply note that the state variables are given by

$$\begin{bmatrix} \mathbf{x}_t \\ \boldsymbol{\lambda}^{\mathbf{y}}_t \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix},$$

and that equation (43) can therefore also be written as

$$\mathbf{v}_{t+1} = \frac{1}{\beta\theta} \begin{bmatrix} \mathbf{C}_1' \left(\mathbf{V}_{11} - \mathbf{V}_{21}' \mathbf{V}_{22}^{-1} \mathbf{V}_{21}\right) & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ -\mathbf{V}_{22}^{-1}\mathbf{V}_{21} & \mathbf{V}_{22}^{-1} \end{bmatrix} \begin{bmatrix} \mathrm{E}_t \mathbf{x}_{t+1} \\ \mathrm{E}_t \boldsymbol{\lambda}^{\mathbf{y}}_{t+1} \end{bmatrix}.$$

instrument(s) to the marginal benefit, where the marginal benefit comes in two forms: a direct effect on the loss function of a change in $\mathrm{E}_t\mathbf{x}_{t+1}$ and an indirect effect through the effect the change in this expectation has on the non-predetermined variables. Third, because its first-order condition holds for all $t > 0$ and does not depend on the non-predetermined variables, the fictitious evil agent is not subject to a time-inconsistency problem. Therefore, the solution to the game in which the Stackelberg leader and the fictitious evil agent both make decisions once-and-for-all is also the solution to the game in which the Stackelberg leader makes decisions once-and-for-all, but the fictitious evil agent makes decisions sequentially.

8 Robust monetary policy

With the solution apparatus behind us, in this section I analyze robust monetary policy in terms of the canonical NK business cycle model (Clarida, Galí, and Gertler, 1999) and the sticky price/sticky wage business cycle model developed by Erceg, Henderson, and Levin (2000). For each model, I formulate the robust Stackelberg game first as per section 3 and then as per section 7. To prevent confusion, when presenting the results I will refer to the optimization problems described in sections 3 and 7 as "problem one" and "problem two," respectively.

8.1 The canonical NK model

The canonical NK business cycle model consists of equations for consumption and aggregate inflation. The model is one in which firms are monopolistically competitive and nominal prices are "sticky" due to Calvo-contracts (Calvo, 1983). Households maximize utility defined over consumption (a Dixit-Stiglitz aggregate of the firms' outputs) and leisure, subject to a budget constraint in which households can transfer income through time by holding either (one-period) nominal bonds or nominal money balances. The labor market is perfectly competitive and labor is the only factor in a constant-returns-to-scale production function. When log-linearized about a zero-inflation steady state, the model has the form

$$c_t = \mathrm{E}_t c_{t+1} - \gamma \left(i_t - \mathrm{E}_t \pi_{t+1} \right) + e_{1t} \tag{44}$$

$$\pi_t = \beta \mathrm{E}_t \pi_{t+1} + \kappa c_t + e_{2t} \tag{45}$$

$$e_{1t} = \rho_1 e_{1t-1} + \sigma_{\varepsilon 1} \varepsilon_{1t} \tag{46}$$

$$e_{2t} = \rho_2 e_{2t-1} + \sigma_{\varepsilon 2} \varepsilon_{2t}, \tag{47}$$

where $\gamma \in (0, \infty)$ is the elasticity of intertemporal substitution, $\beta \in (0, 1)$ is the subjective discount factor, and $\kappa \in (0, \infty)$ represents the slope of the (short-run) Phillips curve and depends on the price rigidity and the discount factor. In the model, c_t denotes consumption, π_t denotes aggregate inflation, and i_t denotes the short-term nominal interest rate. The consumption Euler equation is shifted by a consumption preference shock, e_{1t}, and the Phillips curve is shifted by a markup shock, e_{2t}, with each shock following an AR(1) process with white noise innovations. The parameters in the two shock processes satisfy $\{\rho_1, \rho_2\} \in (-1, 1)$ and $\{\sigma_{\varepsilon 1}, \sigma_{\varepsilon 2}\} \in (0, \infty)$. Subject to equations (44) – (47), the central bank (the Stackelberg leader) chooses $\{i_t\}_0^\infty$ to optimize the loss function,

$$Loss\,[0, \infty] = \mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\pi_t^2 + \lambda c_t^2 + \nu i_t^2\right]. \tag{48}$$

8.1.1 Problem one

When the model is written in state-space form, it is straightforward to solve for the rational expectations equilibrium. By adding the specification errors (the controls of the fictitious evil agent), I can easily obtain the worst-case equilibrium, the approximating equilibria, and the Hansen-Sargent approximating equilibrium. For reasons that will shortly become clear, the robustness parameter, θ, is set to 38.967.[16] Solving for the worst-case equilibrium, the worst-case distortions are

$$v_{t+1}^{\varepsilon_1} = 0.040 e_{1t} + 0.003 e_{2t} - 0.002 p_{\pi t} - 0.005 p_{ct} \tag{49}$$

$$v_{t+1}^{\varepsilon_2} = -0.002 e_{1t} + 0.088 e_{2t} - 0.007 p_{\pi t} + 0.015 p_{ct}, \tag{50}$$

where p_{ct} and $p_{\pi t}$ are the shadow prices associated with consumption and inflation, respectively. Equations (49) and (50) reveal that the distortions that the central bank fears are those that increase the persistence of the shocks, particularly the persistence of the markup shock. These fears are easily rationalized because persistent shocks make it more difficult for the central bank to stabilize the economy. Moreover, persistent markup shocks are especially damaging because they force the central bank to raise interest rates — lowering current-period consumption — in order to unwind the inflationary effects of the shock.

How does the central bank's fear of misspecification affect the approximating equilibrium?

[16] The remaining parameters were assigned the values reported in Giordani and Söderlind (2004). Specifically, I set $\beta = 0.99$, $\gamma = 0.5$, $\kappa = 0.645$, $\rho_1 = \rho_2 = 0.8$, and $\sigma_{\varepsilon 1} = \sigma_{\varepsilon 2} = 1$, with the loss function parameters set to $\lambda = 0.5$ and $\nu = 0.2$.

In the approximating equilibrium, the law of motion for the state variables is

$$\begin{bmatrix} e_{1t+1} \\ e_{2t+1} \\ p_{\pi t+1} \\ p_{ct+1} \end{bmatrix} = \begin{bmatrix} 0.8 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 \\ -0.302 & -0.630 & 0.458 & -0.253 \\ -0.355 & 0.035 & 0.053 & 0.396 \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \\ p_{\pi t} \\ p_{ct} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t+1} \\ \varepsilon_{2t+1} \end{bmatrix}, \quad (51)$$

with inflation, consumption, and the nominal interest rate determined according to

$$\pi_t = 0.124e_{1t} + 0.647e_{2t} + 0.568p_{\pi t} + 0.452p_{ct}, \quad (52)$$

$$c_t = 0.544e_{1t} - 0.904e_{2t} + 0.452p_{\pi t} + 0.633p_{ct}, \quad (53)$$

$$i_t = 0.879e_{1t} - 0.086e_{2t} - 0.132p_{\pi t} - 0.981p_{ct}. \quad (54)$$

By way of contrast, if the approximating equilibrium were constructed using the Hansen-Sargent method, then the law of motion for the state variables would become

$$\begin{bmatrix} e_{1t+1} \\ e_{2t+1} \\ p_{\pi t+1} \\ p_{ct+1} \end{bmatrix} = \begin{bmatrix} 0.8 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 \\ -0.357 & -0.173 & 0.424 & -0.171 \\ -0.278 & -0.414 & 0.086 & 0.312 \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \\ p_{\pi t} \\ p_{ct} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t+1} \\ \varepsilon_{2t+1} \end{bmatrix}, \quad (55)$$

with the behavior of inflation, consumption, and the nominal interest rate still given by equations (52) – (54). Comparing equations (51) and (55), it is clear that the main differences reside in how the shadow prices respond to the shocks, particularly in how they respond to the markup shock.

Figure 1 illustrates how the model, with robust monetary policy, behaves following an adverse 1 percent markup shock, and clarifies why θ was set to 38.967. Panel A maps out the relationship between the (log of the) robustness parameter, θ, and the probability of making a "detection error," given a sample of 200 observations. A detection-error probability is the probability that an econometrician observing equilibrium outcomes would infer incorrectly whether the approximating equilibrium or the worst-case equilibrium generated the data. Among other factors, detection-error probabilities depend on the available sample size. A method for calculating detection-error probabilities is described in Appendix B.[17]

As θ increases (η falls), the distortions become smaller and the probability of making a detection error rises. In fact, in the limit as $\theta \uparrow \infty$ ($\eta \downarrow 0$), the central bank's rising confidence in its model means that the specification errors are so severely constrained that the approximating equilibrium and the worst-case equilibrium each converge to the rational expectations equilibrium, and the detection-error probability converges to 0.5. Panel A shows

[17] See also Hansen, Sargent, and Wang (2002) or Hansen and Sargent (2006, chapter 9).

that using the Hansen-Sargent approximating equilibrium erroneously has no effect on the probability of making a detection error and that choosing $\theta = 38.967$ (so that $\log(\theta) = 3.663$) equates the detection-error probability to 0.10.

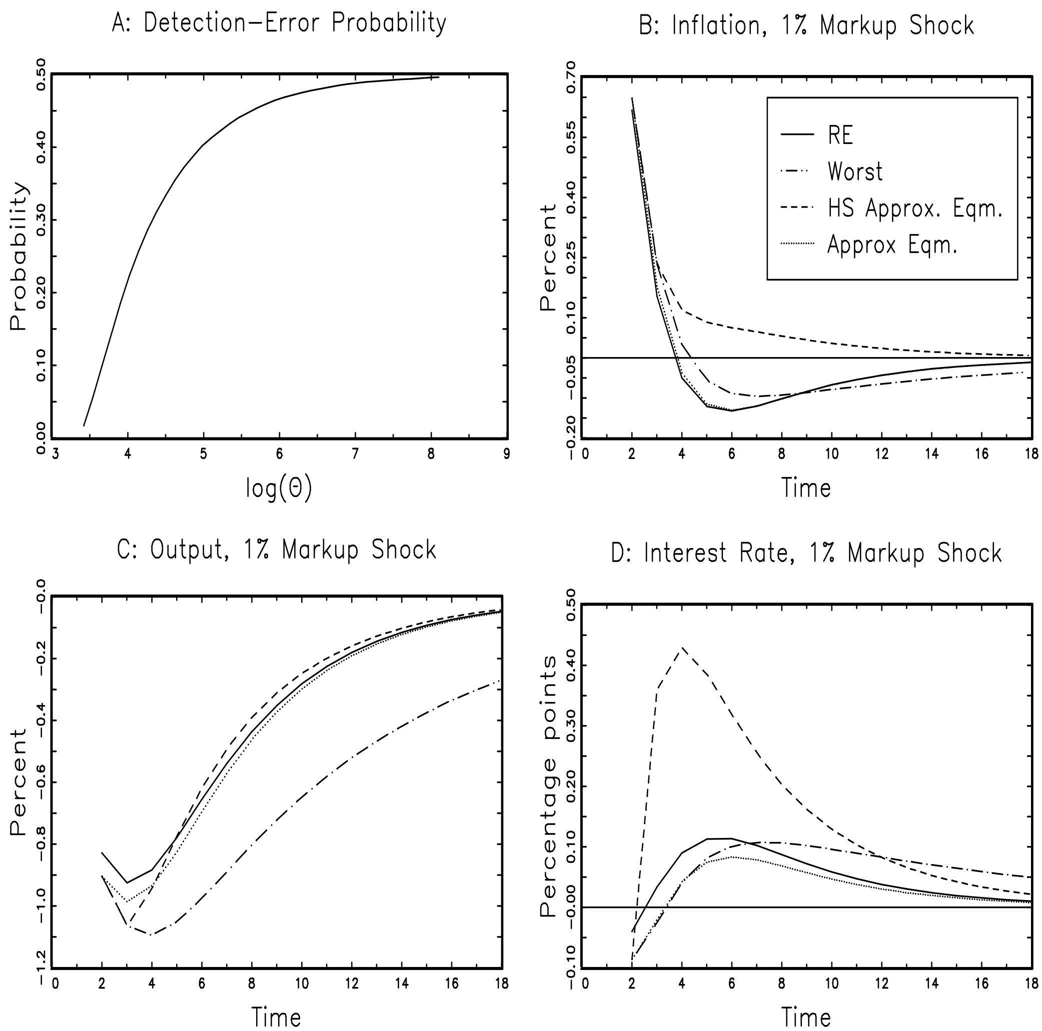

Figure 1: Robust policy in the canonical NK model (problem one)

Although using the Hansen-Sargent approximating equilibrium has no effect on the probability of making a detection error, the remaining panels in Figure 1 demonstrate that its use generates a very different, and surprising, set of dynamics. With $\theta = 38.967$, Panels B – D report the responses of inflation, consumption, and the interest rate to a positive 1 percent markup shock in the rational expectations equilibrium (RE), the worst-case equilibrium (Worst), the Hansen-Sargent approximating equilibrium (HS Approx. Eqm.), and the approximating equilibrium (Approx. Eqm.). In the rational expectations equilibrium, the markup shock causes inflation to rise, with interest rates (after an initial decline) rising in response. The higher interest rate lowers consumption, placing downward pressure on inflation. As inflation begins to decline, the interest rate falls and consumption rises. In the worst-case equilibrium, inflation is more persistent, with the effects of the shock dissipating less rapidly than under rational expectations. To counter this inflation persistence, the central bank commits to a gradual, but sustained, rise in interest rates. In the approximating equilibrium, the economy's initial responses are the same as those for the worst-case equilibrium. Subsequently, however, interest rates rise by less than they do in the worst-case equilibrium and the economy's behavior more closely resembles that of the rational expectations equilibrium.

However, where the responses for the rational expectations, the worst-case, and the approximating equilibria are all relatively similar, the responses for the Hansen-Sargent approximating equilibrium are very different. For example, inflation's response is systematically higher than it is for the other equilibria. In fact, the Hansen-Sargent approximating equilibrium does not show inflation falling below baseline, a hallmark of commitment policies in NK models. With higher inflation, the central bank raises interest rates more, generating a larger decline in consumption. Notice that where the Hansen-Sargent approximating equilibrium suggests that the central bank should respond aggressively to the shock, the approximating equilibrium actually suggests attenuation, with the robust policy responding less aggressively to the shock than under the rational expectations policy.

Finally, an obvious question raised by Figure 1 is why using the Hansen-Sargent approximating equilibrium has no effect on the probability of making a detection error when it has such important implications for the behavior of inflation, consumption, and the interest rate. Setting aside the fact that the probability of making a detection error depends on both shocks, not just the markup shock, it is important to recognize that detection-error probabilities depend on the behavior of the shock processes and not on the behavior of the shadow prices, the decision variables, or the endogenous variables (see Appendix B). As a consequence, the unattractive behavior of inflation, consumption, and the interest rate exhibited in Figure 1 holds no implication for the probability of making a detection error. Clearly, one must be cautious when using detection-error probabilities to measure the "distance" between two models, because behavioral differences that are quantitatively and qualitatively important can be masked.

8.1.2 Problem two

When I formulate the optimization problem as per section 7, setting $\theta = 30.829$ ($\log(\theta) = 3.428$) so that the detection-error probability equals 0.1, the worst-case distortions are

$$v_{t+1}^{\varepsilon_1} = 0.042 e_{1t} + 0.004 e_{2t}, \tag{56}$$

$$v_{t+1}^{\varepsilon_2} = 0.004 e_{1t} + 0.091 e_{2t}. \tag{57}$$

Notice that, unlike problem one, the worst-case distortions depend only on the shocks. Nevertheless, the specification errors that the central bank fears are those that increase the persistence of the markup shock. The law of motion for the state variables in the approximating equilibrium is

$$\begin{bmatrix} e_{1t+1} \\ e_{2t+1} \\ p_{\pi t+1} \\ p_{ct+1} \end{bmatrix} = \begin{bmatrix} 0.8 & 0 & 0 & 0 \\ 0 & 0.8 & 0 & 0 \\ -0.292 & -0.610 & 0.456 & -0.251 \\ -0.346 & 0.016 & 0.054 & 0.393 \end{bmatrix} \begin{bmatrix} e_{1t} \\ e_{2t} \\ p_{\pi t} \\ p_{ct} \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{1t+1} \\ \varepsilon_{2t+1} \end{bmatrix}, \tag{58}$$

with inflation, consumption, and the nominal interest rate determined according to

$$\pi_t = 0.119 e_{1t} + 0.618 e_{2t} + 0.570 p_{\pi t} + 0.448 p_{ct}, \tag{59}$$

$$c_t = 0.531 e_{1t} - 0.829 e_{2t} + 0.448 p_{\pi t} + 0.644 p_{ct}, \tag{60}$$

$$i_t = 0.855 e_{1t} - 0.040 e_{2t} - 0.133 p_{\pi t} - 0.973 p_{ct}. \tag{61}$$

Figure 2 illustrates how the model behaves following an adverse 1 percent markup shock and can be compared directly to Figure 1. Of course, the rational expectations responses are the same in the two figures, however, in the worst-case equilibrium, consumption declines by less in Figure 2 than in Figure 1 and, therefore, the interest rate response in Figure 2 shows the central bank responding more aggressively to the shock than in Figure 1. Turning to the model's behavior in the approximating equilibrium, a striking result in Figure 2 is that the approximating equilibrium is *identical* to the rational expectations equilibrium. In practical terms, the central bank's fear of misspecification and the intensity with which it fears misspecification have no policy implications. In fact, it is straightforward to verify that equations (58) – (61) are the same as rational expectations, implying that the central bank's preference for robustness, because it does not affect policy, does not affect economic outcomes more generally. Strikingly, according to this formulation of the robust control problem, robustness is irrelevant.

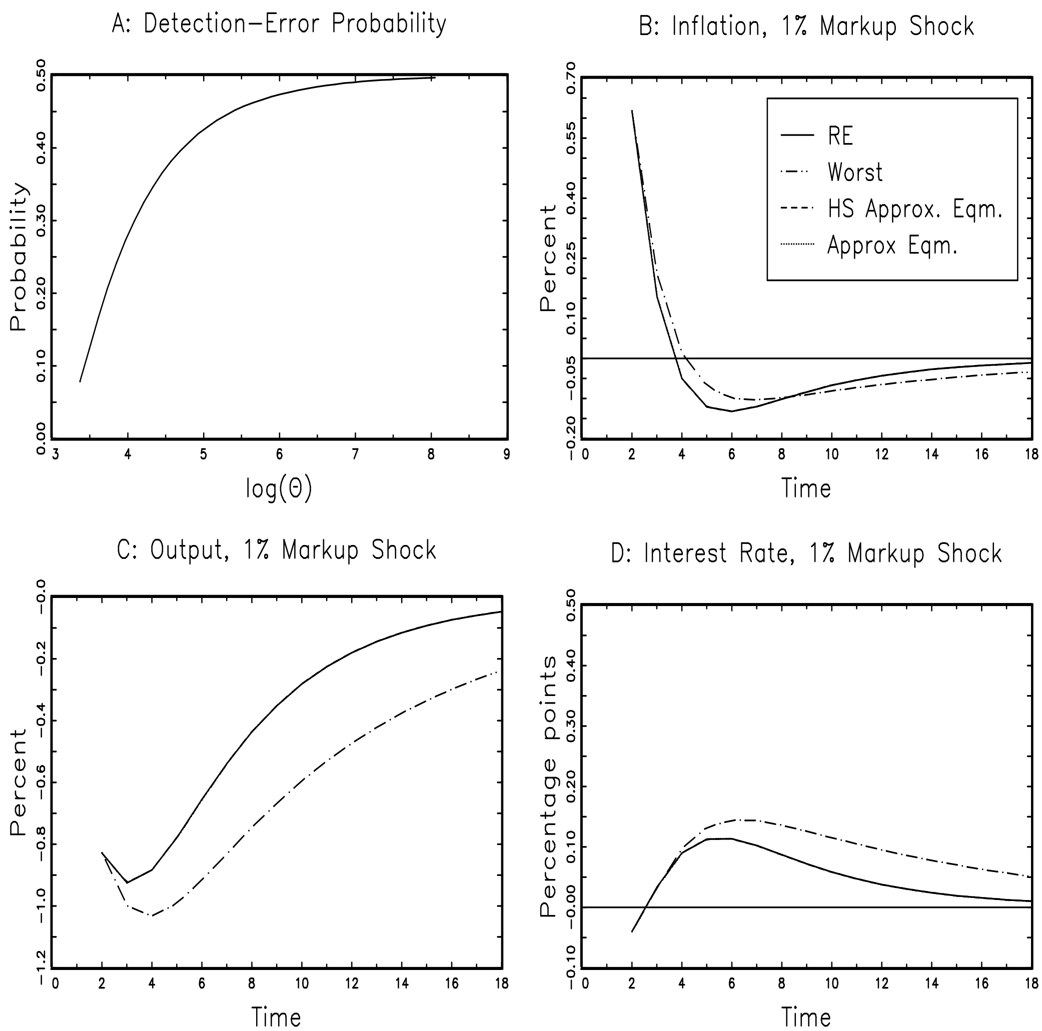

Figure 2: Robust policy in the canonical NK model (problem two)

8.2 A sticky price/sticky wage model

I now turn to the sticky price/sticky wage NK business cycle model developed by Erceg, Henderson, and Levin (2000), a model increasingly reflective of those used to study optimal policymaking. Similar to the canonical NK model, the production technology is linear in labor and prices are subject to a Calvo-style nominal price rigidity. Here, however, nominal wages are also rigid (Calvo wages) with monopolistically competitive firms hiring a Dixit-Stiglitz labor aggregate. In addition to a consumption preference shock, the model contains a leisure preference shock and a serially correlated technology shock.

When log-linearized around its flex-price equilibrium, the model can be written as

$$g_t = \mathrm{E}_t g_{t+1} - \frac{1}{\sigma l_c}\left(i_t - \mathrm{E}_t \pi_{t+1} - r_t^*\right) \tag{62}$$

$$\pi_t = \beta \mathrm{E}_t \pi_{t+1} + \kappa_\rho \left(\overline{w}_t - mpl_t\right) \tag{63}$$

$$w_t = \beta \mathrm{E}_t w_{t+1} + \kappa_w \left(mrs_t - \overline{w}_t\right) \tag{64}$$

$$mpl_t = \overline{w}_t^* - \lambda_{mpl} g_t \tag{65}$$

$$mrs_t = \overline{w}_t^* + \lambda_{mrs} g_t \tag{66}$$

$$\overline{w}_t = \overline{w}_{t-1} + w_t - \pi_t \tag{67}$$

$$r_t^* = \sigma l_c \left(\mathrm{E}_t y_{t+1}^* - y_t^*\right) + \sigma l_q \left(\mathrm{E}_t q_{t+1} - q_t\right) \tag{68}$$

$$\overline{w}_t^* = w_x^* x_t + w_q^* q_t + w_z^* z_t \tag{69}$$

$$y_t^* = y_x^* x_t + y_q^* q_t + y_z^* z_t \tag{70}$$

$$x_t = \rho x_{t-1} + \varepsilon_t, \tag{71}$$

where g_t denotes the output gap, i_t denotes the short-term nominal interest rate, π_t denotes price inflation, $\overline{w}_t$ denotes the real wage, mpl_t denotes the marginal product of labor, w_t denotes wage inflation, mrs_t denotes the marginal rate of substitution between consumption and leisure, and r_t^*, $\overline{w}_t^*$, and y_t^* denote the Pareto-optimal real interest rate, real wage, and level of output, respectively. Finally, x_t, z_t, and q_t represent a technology shock, a leisure preference shock, and a consumption preference shock, respectively.[18]

The central bank's objective is to set the nominal interest rate sequence, $\{i_t\}_0^\infty$, to maximize social welfare, where social welfare is represented by a second-order approximation (around the flex-price equilibrium) of the representative household's utility function. Erceg, Henderson, and Levin (2000) show that in this second-order approximation the expected deviation of social welfare, $\widetilde{W}$, from its Pareto-optimal level, W^*, is equivalent to

$$\begin{aligned} \mathrm{E}\left(W^* - \widetilde{W}\right) \propto\ & \frac{\left(\lambda_{mrs} + \lambda_{mpl}\right)}{2} Var\left(g_t\right) \\ & + \frac{\left(1+\theta_p\right)}{2\theta_p} \frac{\left(1-\beta\xi_p\right)}{\left(1-\xi_p\right)} \frac{1}{\kappa_p} Var\left(\pi_t\right) \\ & + \frac{\left(1+\theta_w\right)}{2\theta_w} \frac{\left(1-\beta\xi_w\right)}{\left(1-\xi_w\right)} \frac{\left(1-\alpha\right)}{\kappa_w} Var\left(w_t\right). \end{aligned} \tag{72}$$

---

[18] I parameterize the model according to Erceg, Henderson, and Levin (2000), setting $\sigma = \chi = 1.5$, $\alpha = 0.3$, $\theta_w = \theta_p = \frac{1}{3}$, $\zeta_w = \zeta_p = 0.75$, $\rho = 0.95$, $\beta = 0.99$, $\overline{C} = 3.163$, $\overline{Q} = 0.3163$, $\overline{N} = 0.27$, $\overline{Z} = 0.03$, $l_c = \frac{\overline{C}}{\overline{C}-\overline{Q}}$, $l_q = \frac{\overline{Q}}{\overline{C}-\overline{Q}}$, $l_n = \frac{\overline{N}}{1-\overline{N}-\overline{Z}}$, $l_z = \frac{\overline{Z}}{1-\overline{N}-\overline{Z}}$, $\Lambda = \alpha + \chi l_n + (1-\alpha)\sigma l_c$, $\lambda_{mpl} = \frac{\alpha}{1-\alpha}$, $\lambda_{mrs} = \sigma l_c + \frac{\chi l_n}{1-\alpha}$, $\kappa_p = \frac{\left(1-\zeta_p\right)\left(1-\beta\zeta_p\right)}{\zeta_p}$, $\kappa_w = \frac{(1-\zeta_w)(1-\beta\zeta_w)}{\left(1+\chi l_n\left(\frac{1+\theta_w}{\theta_w}\right)\right)\zeta_w}$, $w_x^* = \frac{\chi l_n + \alpha l_c}{\Lambda}$, $w_q^* = -\frac{\alpha\sigma l_q}{\Lambda}$, $w_z^* = \frac{\alpha\chi l_z}{\Lambda}$, $y_x^* = \frac{1+\chi l_n}{\Lambda}$, $y_q^* = \frac{(1-\alpha)\sigma l_q}{\Lambda}$, $y_z^* = \frac{-(1-\alpha)\chi l_z}{\Lambda}$.

Because social welfare cannot exceed the Pareto-optimal level, the central bank's optimization problem is to minimize equation (72) subject to equation (7) and equations (62) – (71).

8.2.1 Problem one

As discussed earlier, after adding the specification errors, the only parameter that remains to be determined is the robustness parameter, θ. I set the detection-error probability to 0.1, which implies $\theta = 4.75$ (so that $\log(\theta) = 1.558$), and with this parameterization the specification errors that the central bank fears are

$$v^x_{t+1} = 0.107x_t + 0.001z_t - 0.270\overline{w}_{t-1} - 0.006p_{wt} - 0.001p_{gt} + 0.016p_{\pi t}, \tag{73}$$

$$v^q_{t+1} = -0.001x_t + 0.002\overline{w}_{t-1}, \tag{74}$$

$$v^z_{t+1} = 0.003x_t - 0.006\overline{w}_{t-1} + 0.001p_{\pi t}. \tag{75}$$

Clearly, the central bank is most concerned with specification errors that distort the technology shock, equation (73), and is relatively unconcerned about misspecification of the other shock processes. Because the technology shock plays the same role in this model as the markup shock does in the canonical NK model, this finding is consistent across the two models. Interestingly, the autoregressive coefficient in the worst-case shock process for technology is greater than one. However, stability of the worst-case equilibrium is preserved through feedback among the worst-case shock processes, the real wage, and the shadow prices.

Paralleling Figure 1, Figure 3 (Panel A) plots the relationship between (log) θ and the probability of making a detection error. With $\theta = 4.75$, Figure 3 (Panels B – D) also displays the responses of inflation, output, and the interest rate to a positive 1 percent technology shock for the rational expectations equilibrium, the worst-case equilibrium, and the approximating equilibrium. Focusing first on the rational expectations equilibrium, Figure 3 shows that inflation and the interest rate fall while the output gap rises in response to a 1 percent technological innovation. The positive technology shock lowers the cost of producing a unit of output (real marginal costs), which, through competitive forces, causes prices and inflation (due to sticky prices) to fall. With inflation falling, the central bank cuts interest rates, stimulating demand and opening up a positive output gap. As the effects of the shock begin to dissipate, and as the demand stimulus exerts upward pressure on inflation, both inflation and interest rates begin to rise, causing the output gap to fall. In the worst-case equilibrium, inflation falls by more than under rational expectations and the central bank responds aggressively, cutting interest rates and opening up a larger output gap. Last, the robust policy in the approximating equilibrium initially cuts interest rates by more than when expectations are rational to guard against a sustained and persistent decline in inflation, but then raises interest rates quickly as inflationary pressures begin to build. The central bank's fear of misspecification reveals itself most prominently in how the central bank initially responds to the shock.

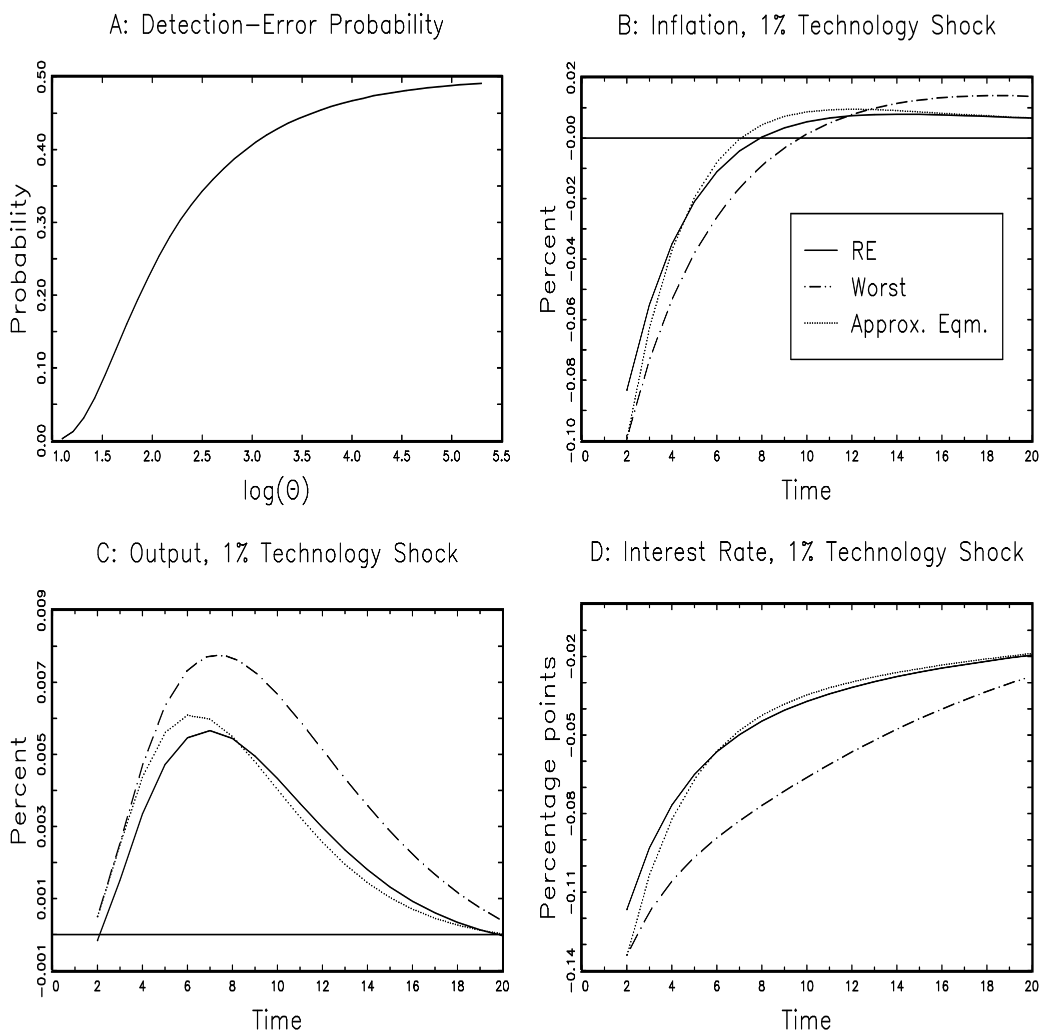

Figure 3: Robust policy in the sticky price/sticky wage model (problem one)

Now consider what happens if the Hansen-Sargent approximating equilibrium is used. Figure 4 reports the responses of inflation, output, and the interest rate, allowing θ to take on three different values. For a reason that will become clear, these three values for θ are all much larger than the value used in Figure 3 and correspond to detection-error probabilities that are all greater than 0.49. The fact that the impulse responses associated with the Hansen-Sargent approximating equilibrium are unusual can be seen immediately in the behavior of the output gap and the interest rate. Where the inflation responses are similar to rational expectations, these inflation paths are associated with an unusual spike in interest rates and the output gap when $\theta = 500$, damped oscillations in interest rates and the output gap when $\theta = 380$, and explosive oscillations in interest rates and the output gap when $\theta = 356$.[19] Notice that these explosive oscillations have no effect on the detection-error probability because the instability does not stem from the shock processes.

Setting the oscillatory behavior aside, the fact that the Hansen-Sargent approximating equilibrium is unstable with $\theta = 356$ is telling, because it demonstrates that the equilibrium does not minimize the maximum loss for distortion sequences that satisfy equation (7). Specifically, the loss associated with the sequence $\{\mathbf{v}_t\}_0^\infty = \{\mathbf{0}\}_0^\infty$ is greater than the loss associated with the sequence of worst-case distortions.

[19] To see the source of this instability, notice that, although the spectral radius of $\mathbf{T}(\mathbf{A} - \mathbf{BF}_u - \mathbf{CF}_v)\mathbf{T}^{-1}$ must be less than one (because the worst-case equilibrium is stable), this does not imply that the spectral radius of $\mathbf{T}(\mathbf{A} - \mathbf{BF}_u)\mathbf{T}^{-1}$ is less than one.

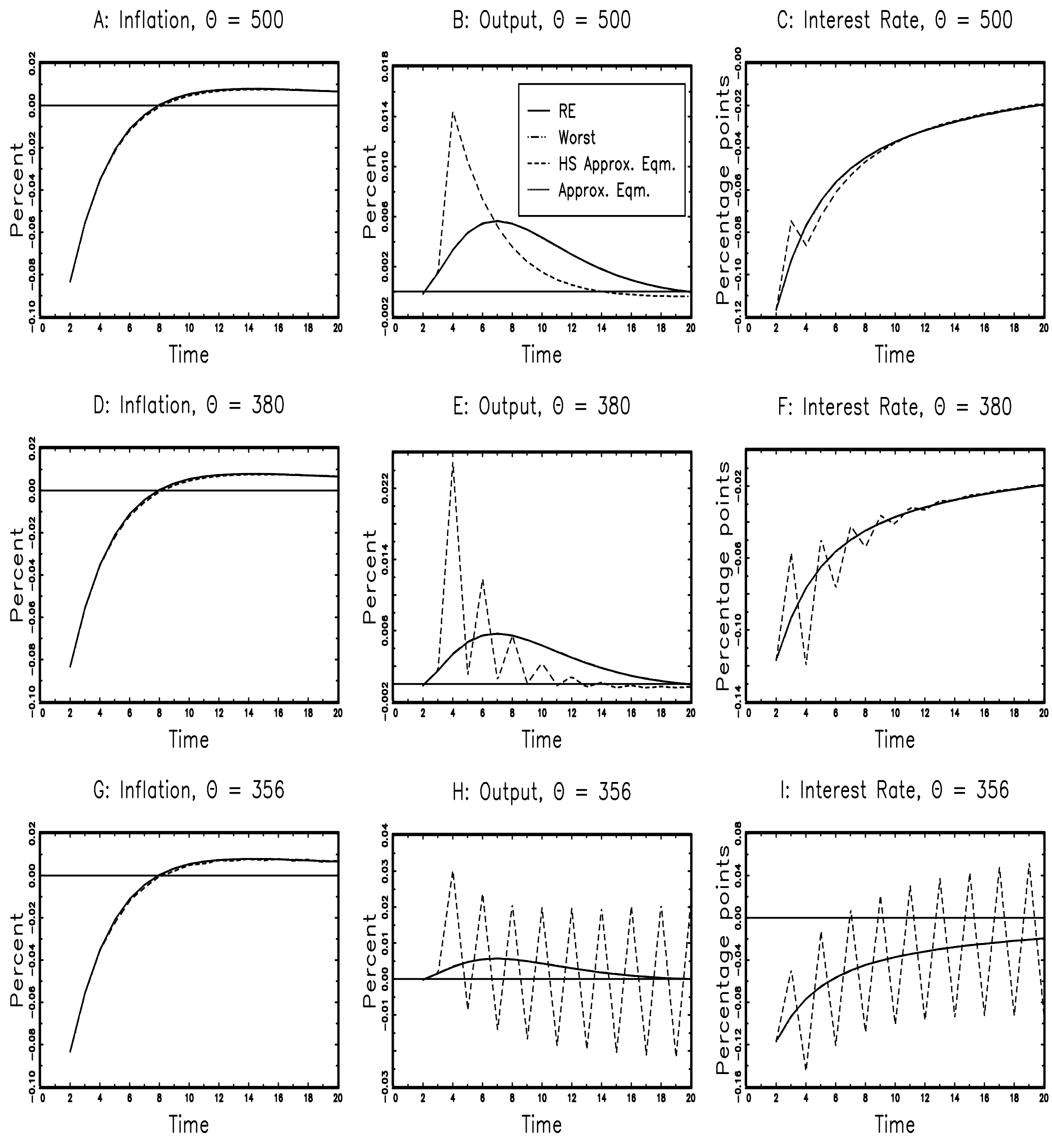

Figure 4: The Hansen-Sargent approximating equilibrium (problem one)

8.2.2 Problem two

I set the detection-error probability to 0.1, which implies $\theta = 3.077$ ($\log(\theta) = 1.124$), and with this parameterization the specification errors that the central bank fears are

$$v_{t+1}^{x} = 0.122x_t + 0.001q_t - 0.003z_t - 0.319\overline{w}_{t-1} - 0.005p_{gt}, \qquad (76)$$

$$v_{t+1}^{q} = -0.001x_t + 0.001\overline{w}_{t-1}, \qquad (77)$$

$$v_{t+1}^{z} = 0.002x_t - 0.004\overline{w}_{t-1}. \qquad (78)$$

As shown earlier, the central bank is most concerned with specification errors that distort the technology shock process, equation (76).

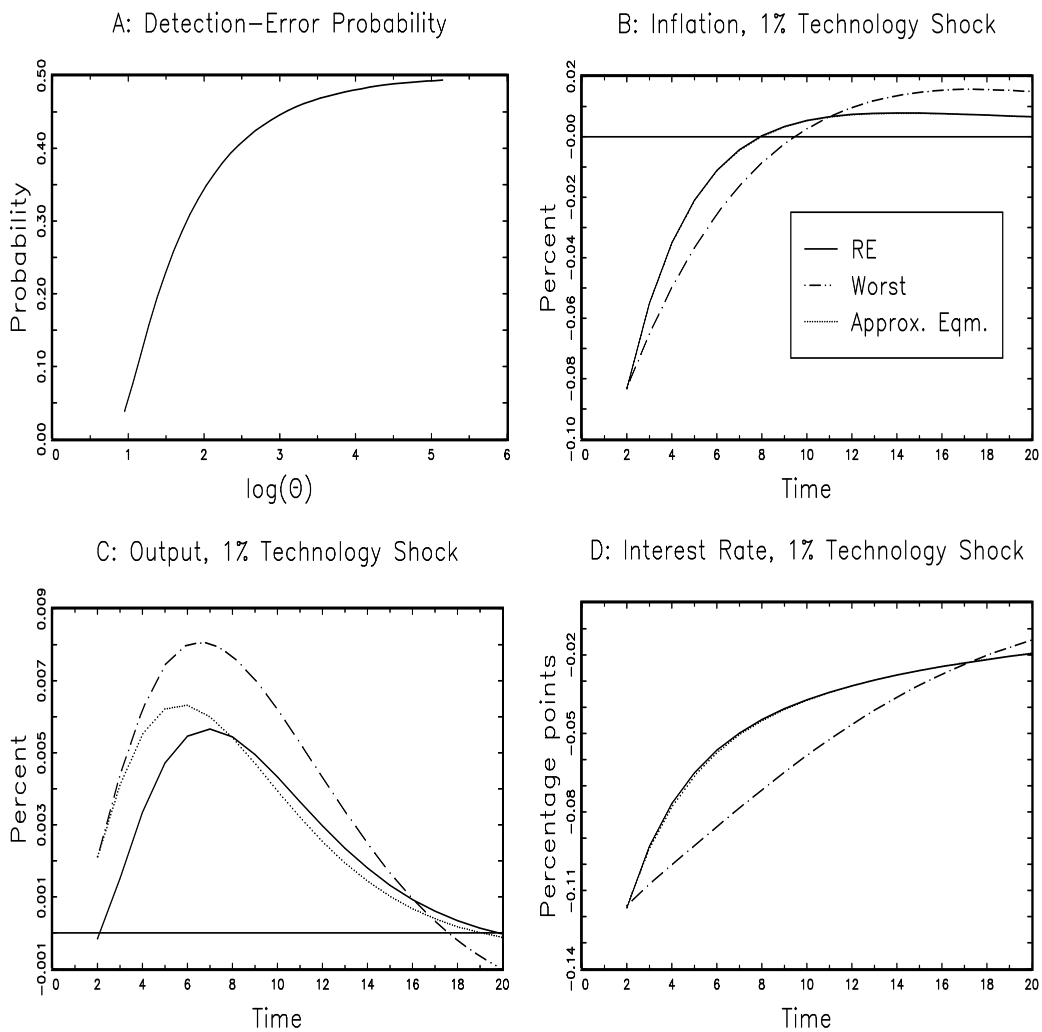

Figure 5: Robust policy in the sticky price/sticky wage model (problem two)

Depicting the economy's response to a technology shock, Figure 5 shows that robustness primarily effects the behavior of output. In this economy, output and consumption are equal, so output is a non-predetermined variable whose behavior is governed partly by central bank promises about future policy. Indeed, comparing the behavior of the robust and the nonrobust economies, the main differences reside in the laws-of-motion for the shadow prices (not shown), particularly those of wage growth and inflation. In effect, although robustness has little effect on the central bank's interest rate decision rule, the central bank's desire for robustness is manifest in the promises it makes about future policy, and these promises assert themselves in the behavior of output. In this respect, Figures 3 and 5 are qualitatively similar. Notably, the interest rate responses in Figure 5 also display a form of attenuation.

9 Conclusion

This paper analyzes decisionmaking in economies where a Stackelberg leader formulates policy under commitment while seeking robustness to model misspecification. The paper's first important contribution is to show that the Hansen and Sargent (2003) solution to this class of robust control problems is generally incorrect. As is by now well-known, shadow prices enter the solution of control problems in which the leader commits. These shadow prices encode the history dependence of the optimal policy and direct private agents toward the equilibrium the Stackelberg leader desires. This paper shows that the approximating equilibrium that Hansen and Sargent (2003) develop is generally incorrect because their method for obtaining it fails to appreciate the role these shadow prices play in determining equilibrium. The paper then shows how the approximating equilibrium can be constructed correctly, establishing as a general principle that the shadow prices must evolve in the approximating equilibrium as they do in the worst-case equilibrium.

Having shown how to construct the approximating equilibrium, this paper analyzes robust monetary policy in two NK business cycle models. An important result that emerges from these applications is that a robust central bank should fear that its approximating model understates the persistence of supply-side shocks, such as markup shocks and technology shocks. Distortions to these shock processes are feared because they are difficult for the central bank to counteract. With the approximating equilibrium constructed correctly, this paper further shows that attenuation does generally characterize robust monetary policy, contrary to popular belief. Further, these applications illustrate that counterintuitive oscillations or even instability can be produced if the Hansen-Sargent approximating equilibrium is used, and that these problems will not be brought to light by detection-error probabilities.

Next the paper considers the treatment of robust Stackelberg games given in Hansen and Sargent (2007), which, due to the issues raised in this paper, differs importantly from Hansen and Sargent (2003). The paper shows that in the solution to this more general formulation of the robust Stackelberg game, conditional on the leader's robust decision rule, the law of motion for the shadow prices is unaffected by the worst-case distortions. As a consequence, the method Hansen and Sargent (2007) use to construct the approximating equilibrium coincides with that developed in this paper because it satisfies the principle that the law of motion for shadow prices must be the same in the approximating equilibrium as it is in the worst-case equilibrium. Further, the paper shows that the first-order condition determining the worst-case distortions does not depend on the non-predetermined variables. Therefore, the worst-case distortions are unaffected by promises about future policy and are also not subject to a time-inconsistency problem.

Returning to the two NK business cycle models, but with the robust Stackelberg game formulated according to Hansen and Sargent (2007), the paper finds that in the canonical NK model the approximating equilibrium coincides with the rational expectations equilibrium for all values of the robustness parameter. Simply put, in this model the central bank's desire for robustness has no implications for monetary policy, for economic outcomes, or for welfare. Fortunately, this stark conclusion does not carry over to the sticky price/sticky wage model, however the main results concerning robust monetary policy do. Specifically, a central bank that seeks robustness should primarily fear specification errors that raise the persistence of supply-side shocks. Finally, although robust control is often associated with aggressive policy responses, for the models analyzed in this paper, and perhaps more generally, attenuation characterizes robust monetary policy.

Appendix A - Worst-case equilibrium

With the optimization problem written as

$$\min_{\{\mathbf{u}_t\}}\max_{\{\mathbf{v}_{t+1}\}} \mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\mathbf{z}_t'\mathbf{W}\mathbf{z}_t + \mathbf{z}_t'\widetilde{\mathbf{U}}\widetilde{\mathbf{u}}_t + \widetilde{\mathbf{u}}_t'\widetilde{\mathbf{U}}'\mathbf{z}_t + \widetilde{\mathbf{u}}_t'\widetilde{\mathbf{R}}\widetilde{\mathbf{u}}_t\right],$$

subject to

$$\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t + \widetilde{\mathbf{B}}\widetilde{\mathbf{u}}_t + \widetilde{\mathbf{C}}\boldsymbol{\varepsilon}_{t+1}, \tag{A1}$$

the policy problem conforms to the standard linear-quadratic framework studied and solved by Backus and Driffill (1986). To obtain the optimal policy, Backus and Driffill (1986) first take $\mathbf{z}_0$ as given and write the optimization problem recursively as

$$\mathbf{z}_t'\mathbf{V}\mathbf{z}_t + d \equiv \min_{\mathbf{u}_t}\max_{\mathbf{v}_{t+1}} \left[\mathbf{z}_t'\mathbf{W}\mathbf{z}_t + \mathbf{z}_t'\widetilde{\mathbf{U}}\widetilde{\mathbf{u}}_t + \widetilde{\mathbf{u}}_t'\widetilde{\mathbf{U}}'\mathbf{z}_t + \widetilde{\mathbf{u}}_t'\widetilde{\mathbf{R}}\widetilde{\mathbf{u}}_t + \beta\mathrm{E}_t\left(\mathbf{z}_{t+1}'\mathbf{V}\mathbf{z}_{t+1} + d\right)\right], \tag{A2}$$

subject to

$$\mathbf{z}_{t+1} = \mathbf{A}\mathbf{z}_t + \widetilde{\mathbf{B}}\widetilde{\mathbf{u}}_t + \widetilde{\mathbf{C}}\boldsymbol{\varepsilon}_{t+1}. \tag{A3}$$

Using dynamic programming they obtain the standard solution

$$\mathbf{z}_{t+1} = (\mathbf{A} - \widetilde{\mathbf{B}}\mathbf{F})\mathbf{z}_t + \widetilde{\mathbf{C}}\boldsymbol{\varepsilon}_{t+1} \tag{A4}$$

$$\widetilde{\mathbf{u}}_t = -\mathbf{F}\mathbf{z}_t, \tag{A5}$$

where

$$\mathbf{F} = \left(\beta\widetilde{\mathbf{B}}'\mathbf{V}\widetilde{\mathbf{B}} + \widetilde{\mathbf{R}}\right)^{-1}\left(\beta\widetilde{\mathbf{B}}'\mathbf{V}\mathbf{A} + \widetilde{\mathbf{U}}\right) \tag{A6}$$

$$\mathbf{V} = \beta\left(\mathbf{A} - \widetilde{\mathbf{B}}\mathbf{F}\right)'\mathbf{V}\left(\mathbf{A} - \widetilde{\mathbf{B}}\mathbf{F}\right) + \mathbf{W} - \widetilde{\mathbf{U}}\mathbf{F} - \widetilde{\mathbf{U}}'\mathbf{F}' + \mathbf{F}'\widetilde{\mathbf{R}}\mathbf{F} \tag{A7}$$

$$d = \frac{\beta tr\left[\widetilde{\mathbf{C}}'\mathbf{V}\widetilde{\mathbf{C}}\mathrm{E}\left(\varepsilon_{t+1}\varepsilon'_{t+1}\right)\right]}{1-\beta}. \tag{A8}$$

Since $\mathbf{y}_0$ is not predetermined and the expectation errors are not exogenous, the Stackelberg leader also chooses $\mathbf{y}_0$ and $\{\varepsilon_{yt}\}_1^\infty$ to minimize the value function

$$\mathbf{z}_0'\mathbf{V}\mathbf{z}_0 + d = \begin{bmatrix} \mathbf{x}_0' & \mathbf{y}_0' \end{bmatrix}\begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{x}_0 \\ \mathbf{y}_0 \end{bmatrix} + d, \tag{A9}$$

obtaining

$$\mathbf{y}_0 = -\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{x}_0 \tag{A10}$$

$$\varepsilon_{yt} = -\mathbf{V}_{22}^{-1}\mathbf{V}_{21}\mathbf{C}_1\varepsilon_{xt}. \tag{A11}$$

In the equilibrium that the leader desires, in the initial period the non-predetermined variables relate to the predetermined variables according to equation (A10) and in each period the expectational errors relate to the innovations according to equation (A11). To achieve this equilibrium, the leader represents its policy as a function of a vector of shadow prices of the non-predetermined variables with the shadow prices inducing private agents to form expectations that are consistent with the desired equilibrium. Thus, introducing a vector of shadow prices of the non-predetermined variables

$$\mathbf{p}_t \equiv \begin{bmatrix} \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix} \tag{A12}$$

and noting that $\mathbf{p}_t$ is predetermined (implied by equation (A11)) with initial value $\mathbf{p}_0 = \mathbf{0}$ (implied by equation (A10)), the dynamics of $\mathbf{x}_t$ and $\mathbf{p}_t$ are given by

$$\begin{bmatrix} \mathbf{x}_{t+1} \\ \mathbf{p}_{t+1} \end{bmatrix} = \mathbf{T}\left(\mathbf{A} - \widetilde{\mathbf{B}}\mathbf{F}\right)\mathbf{T}^{-1}\begin{bmatrix} \mathbf{x}_t \\ \mathbf{p}_t \end{bmatrix} + \mathbf{T}\widetilde{\mathbf{C}}\varepsilon_{t+1} \tag{A13}$$

$$= \mathbf{T}\left(\mathbf{A} - \widetilde{\mathbf{B}}\mathbf{F}\right)\mathbf{T}^{-1}\begin{bmatrix} \mathbf{x}_t \\ \mathbf{p}_t \end{bmatrix} + \begin{bmatrix} \mathbf{C}_1 \\ \mathbf{0} \end{bmatrix}\varepsilon_{xt+1} \tag{A14}$$

where

$$\mathbf{T} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}. \tag{A15}$$

Finally, $\mathbf{y}_t$ and $\widetilde{\mathbf{u}}_t$ are determined according to

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{V}_{22}^{-1}\mathbf{V}_{21} & \mathbf{V}_{22}^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{x}_t \\ \mathbf{p}_t \end{bmatrix} \tag{A16}$$

and

$$\widetilde{\mathbf{u}}_t = -\mathbf{F}\mathbf{T}^{-1}\begin{bmatrix} \mathbf{x}_t \\ \mathbf{p}_t \end{bmatrix}, \tag{A17}$$

respectively.

Equations (A14), (A16), and (A17) describe the state-contingent behavior of the predetermined variables, $\mathbf{x}_t$ and $\mathbf{p}_t$, the non-predetermined variables, $\mathbf{y}_t$, and the decision variables, $\widetilde{\mathbf{u}}_t$, in the unique stable worst-case equilibrium in which the leader sets policy with commitment.

Appendix B - Detection-error probability

Let A denote the approximating model and B denote the worst-case model; then, assigning equal prior weight to each model and assuming that model selection is based on the likelihood ratio principle, Hansen, Sargent, and Wang (2002) show that detection-error probabilities are calculated according to

$$p(\theta) = \frac{\text{prob}(A|B) + \text{prob}(B|A)}{2}, \tag{B1}$$

where prob$(A|B)$ (prob$(B|A)$) represents the probability that the econometrician erroneously chooses A (B) when B (A) generated the data. Let $\{\mathbf{z}_t^B\}_1^T$ denote a finite sequence of economic outcomes (the shocks, the shadow prices, the endogenous variables, and the followers' and leader's decision variables) generated by the worst-case equilibrium, and let L_{AB} and L_{BB} denote the likelihood associated with models A and B, respectively; then the econometrician chooses A over B if $\log(L_{BB}/L_{AB}) < 0$. Generating M independent sequences $\{\mathbf{z}_t^B\}_1^T$, prob$(A|B)$ can be calculated according to

$$\text{prob}(A|B) \approx \frac{1}{M}\sum_{m=1}^{M} \text{I}\left[\log\left(\frac{L_{BB}^m}{L_{AB}^m}\right) < 0\right], \tag{B2}$$

where $\text{I}[\log(L_{BB}^m/L_{AB}^m) < 0]$ is the indicator function that equals one when its argument is satisfied and equals zero otherwise; prob$(B|A)$ is calculated analogously using data generated from the approximating model.

Let

$$\mathbf{z}_{t+1} = \mathbf{H}_A\mathbf{z}_t + \mathbf{G}\boldsymbol{\varepsilon}_{t+1} \tag{B3}$$

$$\mathbf{z}_{t+1} = \mathbf{H}_B\mathbf{z}_t + \mathbf{G}\boldsymbol{\varepsilon}_{t+1} \tag{B4}$$

govern equilibrium outcomes under the approximating equilibrium and the worst-case equilibrium, respectively. Using the Moore-Penrose inverse,

$$\widehat{\boldsymbol{\varepsilon}}_{t+1}^{i|j} = \left(\mathbf{G}'\mathbf{G}\right)^{-1}\mathbf{G}'\left(\mathbf{z}_{t+1}^j - \mathbf{H}_i\mathbf{z}_t^j\right), \quad \{i,j\} \in \{A,B\} \tag{B5}$$

are the inferred innovations in period $t+1$ when model i is fitted to data $\{\mathbf{z}_t^j\}_1^T$ generated from model j, and let $\widehat{\boldsymbol{\Sigma}}^{i|j}$ be the associated estimates of the innovation variance-covariance matrices. Note that the Moore-Penrose inverse picks out the shock process from among the variables in $\mathbf{z}_t$.

Assuming that the innovations are normally distributed, it is easy to show that

$$\log\left(\frac{L_{AA}}{L_{BA}}\right) = \frac{1}{2}\text{tr}\left(\widehat{\boldsymbol{\Sigma}}^{B|A} - \widehat{\boldsymbol{\Sigma}}^{A|A}\right) \tag{B6}$$

$$\log\left(\frac{L_{BB}}{L_{AB}}\right) = \frac{1}{2}\text{tr}\left(\widehat{\boldsymbol{\Sigma}}^{A|B} - \widehat{\boldsymbol{\Sigma}}^{B|B}\right). \tag{B7}$$

Given equations (B6) and (B7), equation (B2) is used to estimate prob$(A|B)$ and (similarly) prob$(B|A)$, which are needed to construct the detection-error probability, as per equation (B1). The robustness parameter, θ, is then determined by selecting a detection-error probability and inverting equation (B1). This inversion can be performed numerically by constructing the mapping between θ and the detection-error probability for a given sample size. Note that any sequence of specification errors that satisfies equation (7) will be at least as difficult to distinguish from the approximating model as is a sequence that satisfies equation (7) with equality; as such, $p(\theta)$ represents a lower bound on the probability of making a detection error.

References

[1] Backus, D., and J. Driffill, (1986), "The Consistency of Optimal Policy in Stochastic Rational Expectations Models," Centre for Economic Policy Research Discussion Paper #124.

[2] Cagetti, M., Hansen, L., Sargent, T., and N. Williams, (2002), "Robustness and Pricing with Uncertain Growth," *The Review of Financial Studies*, 15, 2, pp363-404.

[3] Calvo, G., (1983), "Staggered Contracts in a Utility-Maximising Framework," *Journal of Monetary Economics*, 12, pp383-398.

[4] Clarida, R., J. Galí and M. Gertler, (1999), "The Science of Monetary Policy: A New Keynesian Perspective," *Journal of Economic Literature*, 37, 4, pp1661-1707.

[5] Currie, D., and P. Levine, (1985), "Macroeconomic policy design in an interdependent world," in Buiter, W. and R. Marston (eds) *International Economic Policy Coordination*, Cambridge University Press, Cambridge.

[6] Currie, D., and P. Levine, (1993), *Rules, Reputation and Macroeconomic Policy Coordination*, Cambridge University Press, Cambridge.

[7] Dennis, R., Leitemo, K., and U. Söderström, (2006a), "Methods for Robust Control," Federal Reserve Bank of San Francisco Working Paper # 2006-10.

[8] Dennis, R., Leitemo, K., and U. Söderström, (2006b), "Monetary Policy in a Small Open Economy with a Preference for Robustness," Federal Reserve Bank of San Francisco, mimeo.

[9] Epstein, L., and T. Wang, (1994), "Intertemporal Asset Pricing Under Knightian Uncertainty," *Econometrica*, 62, 3, pp283-322.

[10] Erceg, C., Henderson, D., and A. Levin, (2000), "Optimal Monetary Policy with Staggered Wage and Price Contracts," *Journal of Monetary Economics*, 46, pp281-313.

[11] Giordani, P., and P. Söderlind, (2004), "Solution of Macro-Models with Hansen-Sargent Robust Policies: Some Extensions," *Journal of Economic Dynamics and Control*, 28, pp2367-2397.

[12] Greenspan, A., (2003), Opening Remarks at "Monetary Policy Under Uncertainty," symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming.

[13] Gul, F. (1991), "A Theory of Disappointment Aversion," *Econometrica*, 59, 3, pp667-686.

[14] Hansen, L., and T. Sargent, (2001a), "Acknowledging Misspecification in Macroeconomic Theory," *Review of Economic Dynamics*, 4, pp519-535.

[15] Hansen, L., and T. Sargent, (2001b), "Robust Control and Model Uncertainty," *American Economic Review, Papers and Proceedings*, 91, 2, pp60-66.

[16] Hansen, L., and T. Sargent, (2003), "Robust Control of Forward-Looking Models," *Journal of Monetary Economics*, 50, pp581-604.

[17] Hansen, L., and T. Sargent, (2007), *Robustness*, Princeton University Press, forthcoming, (version dated December 5, 2006).

[18] Hansen, L., Sargent, T., and T. Tallarini, (1999), "Robust Permanent Income and Pricing," *The Review of Economic Studies*, 66, 4, pp873-907.

[19] Hansen, L., Sargent, T., and N. Wang, (2002), "Robust Permanent Income and Pricing with Filtering," *Macroeconomic Dynamics*, 6, pp40-84.

[20] Kwakernaak, H., and R. Sivan, (1972), *Linear Optimal Control Systems*,Wiley Press, New York.

[21] Kydland, F., and E. Prescott, (1980), "Dynamic Optimal Taxation, Rational Expectations and Optimal Control," *Journal of Economic Dynamics and Control*, 2, pp79-91.

[22] Luenberger, D., (1969), *Optimization by Vector Space Methods*, Wiley Press, New York.

[23] Marcet, A., and R. Marimon, (1999), "Recursive Contracts," University of Pompeu Fabra, mimeo.

[24] Oudiz, G., and J. Sachs, (1985), "International policy coordination in dynamic macroeconomic models," in Buiter, W. and R. Marston (eds) *International Economic Policy Coordination*, Cambridge University Press, Cambridge, pp275-319.

[25] Routledge, B., and S. Zin, (2003), "Generalized Disappointment Aversion and Asset prices," National Bureau of Economic Research Working Paper #10107.

[26] Sargent, T., (1999), "Comment: Policy Rules for Open Economies," in Taylor, J. (ed), *Monetary Policy Rules*, University of Chicago Press, Chicago.

[27] Whittle, P., (1990), *Risk Sensitive Optimal Control,* Wiley Press, New York.

[28] Woodford, M., (2005), "Robustly Optimal Monetary Policy with Near-Rational Expectations," University of Columbia, mimeo, (version dated December 13, 2005).