DFT705.tex

Monetary Policy with Model Uncertainty: Distribution Forecast Targeting*

Lars E.O. Svensson and Noah Williams
www.princeton.edu/svensson www.princeton.edu/∼noahw
Princeton University

First draft: May 2005
This version: May 2007

Abstract

We examine optimal and other monetary policies in a linear-quadratic setup with a relatively general form of model uncertainty, so-called Markov jump-linear-quadratic systems extended to include forward-looking variables and unobservable "modes." The form of model uncertainty our framework encompasses includes: simple i.i.d. model deviations; serially correlated model deviations; estimable regime-switching models; more complex structural uncertainty about very different models, for instance, backward- and forward-looking models; time-varying central-bank judgment about the state of model uncertainty; and so forth. We provide an algorithm for finding the optimal policy as well as solutions for arbitrary policy functions. This allows us to compute and plot consistent distribution forecasts—fan charts—of target variables and instruments. Our methods hence extend certainty equivalence and "mean forecast targeting" to more general certainty non-equivalence and "distribution forecast targeting."

JEL Classification: E42, E52, E58
Keywords: Optimal policy, multiplicative uncertainty

*Satoru Shimizu provided excellent research assistance. We thank Pierpaolo Benigno, Alessandro Flamini, Marvin Goodfriend, Boris Hoffman, Eric Leeper, Fabio Milani, Alexei Onatski, Rujikorn Pavasuthipaisit, and participants in the New York Area Workshop on Monetary Policy, New York, May 2005, the Conference on Macroeconomic Risk and Policy Responses, Berlin, May 2005, the NBER Monetary Economics program meeting, Cambridge, April 2006, and seminar audiences at the University of Iowa, Rutgers University, UC-Irvine, Yale University, Sveriges Riksbank, and the Federal Reserve Banks of Philadelphia and Richmond for helpful comments. Expressed views and any remaining errors are our own responsibility.

Contents

1 Introduction

In recent years there has been a renewed interest in the study of optimal monetary policy under uncertainty. Typical formulations of optimal policy consider only additive sources of uncertainty, where in a linear-quadratic (LQ) framework the well-known certainty-equivalence result applies and implies that optimal policy is the same as if there were no uncertainty. Recognizing the uncertain environment that policymakers face, recent research has considered broader forms of uncertainty for which certainty equivalence no longer applies. While this may have important implications, in practice the design of policy becomes much more difficult outside the classical LQ framework.

One of the conclusions of the Onatski and Williams [31] study of model uncertainty is that, for progress to be made, the structure of the model uncertainty has to be explicitly modeled. In line with this, in this paper we develop a very explicit but still relatively general form of model uncertainty that remains quite tractable. We use a so-called Markov jump-linear-quadratic (MJLQ) model, where model uncertainty takes the form of different "modes" (or regimes) that follow a Markov process. Our approach allows us to move beyond the classical linear-quadratic world with additive shocks, yet remains close enough to the LQ framework that the analysis is transparent. We examine optimal and other monetary policies in an extended linear-quadratic setup, extended in a way to capture model uncertainty. The forms of model uncertainty our framework encompasses include: simple i.i.d. model deviations; serially correlated model deviations; estimable regime-switching models; more complex structural uncertainty about very different models, for instance, backward- and forward-looking models; time-varying central-bank judgment—information, knowledge, and views outside the scope of a particular model (Svensson [39])—about the state of model uncertainty; and so forth. Moreover, while we focus on model uncertainty, our methods also apply to other linear models with changes of regime which may capture boom/bust cycles, productivity slowdowns and accelerations, switches in monetary and/or fiscal policy regimes, and so forth. We provide an algorithm for finding the optimal policy as well as solutions for arbitrary policy functions. This allows us to compute and plot consistent distribution forecasts—fan charts—of target variables and instruments. Our methods hence extend certainty equivalence and "mean forecast targeting," where only the mean of future variables matter (Svensson [39]), to more general certainty non-equivalence and "distribution forecast targeting," where the whole probability distribution of future variables matter (Svensson [38]).[1]

[1] The importance of the whole distribution of future target variables was recently emphasized by Greenspan [20] at the 2005 Jackson Hole symposium, with reference to his [19] so-called risk-management approach:

> In this [risk management] approach, a central bank needs to consider not only the most likely [rather: mean] future path for the economy but also the distribution of possible outcomes about that path. The decisionmakers then need to reach a judgment about the probabilities, costs, and benefits of various possible outcomes under alternative choices for policy.

We agree with Feldstein [17] that Greenspan's risk-management approach is best interpreted as standard expected-loss minimization and we consider the risk-management approach and the approach of this paper as completely consistent. See Blinder and Reis [5] for further discussion of possible interpretations of the risk-management approach.

Certain aspects of our approach have been known in economics since the classic works of Aoki [2] and Chow [8], who allowed for multiplicative uncertainty in a linear-quadratic framework. The insight of those papers, when adapted to our setting, is that in MJLQ models the value function for the optimal policy design problem remains quadratic in the state, but now with weights that depend on the mode. MJLQ models have also been widely studied in the control-theory literature for the special case when there are no forward-looking variables (see Costa and Fragoso [10], Costa, Fragoso, and Marques [11] (henceforth CFM), do Val, Geromel, and Costa [15], and the references therein). More recently, Zampolli [45] uses an MJLQ model to examine monetary policy under shifts between regimes with and without an asset-market bubble, although still in a model without forward-looking variables. Blake and Zampolli [4] provide an extension of the MJLQ model to include forward-looking variables, although with less generality than in our paper and with the analysis and the algorithms restricted to observable modes and discretion equilibria.

Our MJLQ approach is also closely related to the Markov regime-switching models which have been widely used in empirical work. These methods first gained prominence with Hamilton [21] which started a burgeoning line of research. Models of this type have been used to study a host empirical phenomena, with many developments and techniques. summarized in [25]. More recently, the implications of Markov switching in rational expectations models of monetary policy have been studied by Davig and Leeper [13] and Farmer, Waggoner, and Zha [16]. These papers focus on (and debate) the conditions for uniqueness or indeterminacy of equilibria in forward-looking models, taking as given a specified policy rule.

Relative to this previous literature, one main contribution of our paper is the development of a general approach for solving for the optimal policy in MJLQ models that include forward-looking variables. This extension is key for policy analysis under rational expectations, but the forward-looking variables make the model nonrecursive. We show that the recursive saddlepoint method of Marcet and Marimon [29] can nevertheless be applied to express the model in a convenient recursive way, and we derive an algorithm for determining the optimal policy and value functions.

A second main contribution of our paper is to deal with the case of unobservable modes. The existing literature has almost exclusively focused on the case where agents can directly observe the mode. While this may be plausible for some environments, such as for example when a new policy regime is announced, in many cases it is more fitting to assume that the modes are not observable. When the modes are not observable, we can represent the decision maker's information as a probability distribution over possible modes, and optimal policy will depend on that distribution. In this paper, we analyze the special case where decision makers do not learn from observations of the economy, but rather the future subjective distribution over modes is entirely governed by the transition probabilities. In this case, the value function remains quadratic in the state, but with weights that depend now on the probability distribution over modes. We develop algorithms for solving this case.

In addition to considering the optimal policy, we also consider the behavior of the model for arbitrary time-varying or time-invariant instrument rules. This allows us to construct model-consistent probability distributions—fan charts—of the variables relevant to policy makers for any arbitrary instrument-rate path. Moreover, much of the literature in monetary policy analysis has focused on "simple" instrument rules which are restricted to respond to only a subset of all available information, with Taylor rules and various generalizations being most prominent. We show how to derive optimal restricted instrument rules in our setting. Importantly, our approach is not restricted to *instrument* rules; *any* given or optimal restricted policy rule, including *targeting* rules, can be considered.

The more general case where modes are unobservable and decision makers infer from their observations the probability of being in a particular mode is much more difficult to solve. The optimal filter is nonlinear, which destroys the tractability of the MJLQ approach.[2] Additionally, as in most Bayesian learning problems, the optimal policy will also include an experimentation component. Thus, solving for the optimal decision rules will be a more complex numerical task. Due to the curse of dimensionality, it is only feasible in models with a relatively small number of state variables and modes. Confronted with these difficulties, the literature has focused on approximations such as linearization or adaptive control.[3] While these issues are important, they remain outside the scope of the present paper and are instead examined in Svensson and Williams [42].

The rest of the paper is organized as follows. The bulk of our analysis is carried out in section 2. There we lay out the model, discuss the differing informational assumptions we employ, and show how to solve for the optimal policy. In section 3, we discuss how different kinds of model uncertainty can be incorporated by our framework. In section 4, we present examples based on two empirical models of the US economy: regime-switching versions of the backward-looking model of Rudebusch and Svensson [33] and the forward-looking New Keynesian model of Lindé [27]. We also include one estimated example focusing directly on the role of private sector expectations in policy choice. In section 5, we show how the same probability distributions can be constructed for arbitrary time-invariant instrument rules and optimal restricted instrument rules. Here we derive optimal generalized and mode-dependent Taylor-type rules in the Lindé model. In section 6, we present some conclusions. The appendices contain some technical details and extensions of the material in the text.

[2] The optimal nonlinear filter is well-known, and it is a key component of the estimation methods as well (Hamilton [22] and Kim and Nelson [25]).

[3] In the first case, restricting attention to (sub-optimal) linear filters preserves the tractability of the linear-quadratic framework. See CFM [11] for a brief discussion and references. In adaptive control, agents do not take into account the informational role of their decisions. See do Val and Başar [14] for an application of an adaptive control MJLQ problem in economics. In a different setting, Cogley, Colacito, and Sargent [9] have recently studied how well adaptive procedures approximate the optimal policies.

2 The model

We set up a relatively flexible model of an economy with a private sector and a central bank, which allows for relatively broad additive and multiplicative uncertainty as well as different relevant representations of the central-bank information and judgment about the economy.

2.1 The baseline model

As our benchmark, we consider the following model of an economy with a central bank:

$$X_{t+1} = A_{11j_{t+1}} X_t + A_{12j_{t+1}} x_t + B_{1j_{t+1}} i_t + C_{j_{t+1}} \varepsilon_{t+1}, \tag{2.1}$$

$$\mathrm{E}_t H_{j_{t+1}} x_{t+1} = A_{21j_t} X_t + A_{22j_t} x_t + B_{2j_t} i_t, \tag{2.2}$$

where X_t is an n_X-vector of predetermined variables (the state) in period t (the first element may be unity to incorporate nonzero intercepts in a convenient way), x_t is an n_x-vector of forward-looking variables in period t, i_t is an n_i-vector of central-bank instruments (control variables) in period t, and ε_t is an n_X-vector of zero-mean i.i.d. shocks realized in period t with covariance matrix I_{n_X}. The forward-looking variables and the instruments are the nonpredetermined variables.[4]

The matrices A_{11j_t}, A_{12j_t}, B_{1j_t}, C_{j_t}, H_{j_t}, A_{21j_t}, A_{22j_t}, and B_{2j_t} (assumed to be of appropriate dimension) are random and can each take n_j different values in period t, corresponding to the n_j *modes* $j_t \in N_j \equiv \{1, 2, ..., n_j\}$ in period t. The modes j_t follow a Markov process with constant transition probabilities:

$$P_{jk} \equiv \Pr\{j_{t+1} = k \,|\, j_t = j\} \qquad (j, k \in N_j). \tag{2.3}$$

While we focus throughout on the time-homogeneous case, it is straightforward to allow the modes to depend directly on calendar time. Furthermore, P denotes the $n_j \times n_j$ transition matrix $[P_{jk}]$ and the n_j-vector $p \equiv (p_{1t}, ..., p_{n_j t})'$ (where $p_{jt} \equiv \Pr\{j_t = j\}$, $j \in N_j$) denotes the probability distribution of the modes in period t, so

$$p_{t+1} = P' p_t.$$

Finally, the n_j-vector $\bar{p}$ denotes the unique stationary distribution of the modes, so[5]

$$\bar{p} = P' \bar{p}.$$

We assume that the matrix A_{22j_t} is nonsingular for each $j_t \in N_j$, so equation (2.2) determines the forward-looking variables in period t. There is no restriction in including the shock ε_t only in the

[4] Predetermined variables have exogenous one-period-ahead forecast errors, whereas non-predetermined variables have endogenous one-period-ahead forecast errors.

[5] We assume that the Markov chain is recurrent and aperiodic, so the stationary distribution is unique and does not depend on the initial mode (Karlin and Taylor [24]). A simple sufficient condition is that the matrix P^m has all elements positive for some $m \geq 1$ (Ljungqvist and Sargent [28]).

equations for the predetermined variables, since, if necessary, the set of predetermined variables can always be expanded to include the shocks and the shocks this way indirectly enter into the equations for the forward-looking variables. The shocks ε_t and the modes j_t are assumed to be independently distributed (although we allow the impact on the economy of the shocks ε_t to depend on the modes j_t through the matrix C_{j_t}). However, this assumption is not restrictive. Mode-dependent additive shocks are actually incorporated, since the fact that we allow one of the predetermined variables to be unity implies that all our equations may have mode-dependent intercepts.[6] For any random variable q_{t+1} realized in period $t+1$ the expression $\mathrm{E}_t q_{t+1}$ denotes the conditional expectation of the central bank and the private sector in period t; we hence assume that information is symmetric between the central bank and the private sector. The precise informational assumptions underlying the conditional expectations operator E_t are specified below.

The central bank has an intertemporal loss function in period t,

$$\mathrm{E}_t \sum_{\tau=0}^{\infty} \delta^\tau L(X_{t+\tau}, x_{t+\tau}, i_{t+\tau}, j_{t+\tau}), \tag{2.4}$$

where the period loss function, $L(X_t, x_t, i_t, j_t)$, satisfies

$$L(X_t, x_t, i_t, j_t) \equiv Y_t' \Lambda_{j_t} Y_t,$$

where

$$Y_t \equiv D_{j_t} \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix}$$

is an n_Y-vector of target variables and the weight matrix Λ_{j_t} depends on the mode j_t and is symmetric and positive semidefinite for each $j_t \in N_j$. It follows that the period loss function satisfies

$$L(X_t, x_t, i_t, j_t) \equiv \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix}' W_{j_t} \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix}, \tag{2.5}$$

where the matrix $W_{j_t} \equiv D_{j_t}' \Lambda_{j_t} D_{j_t}$ depends on the mode j_t and is symmetric and positive semidefinite for each $j_t \in N_j$. The scalar δ is a discount factor satisfying $0 < \delta \leq 1$.[7]

2.2 Informational assumptions

We consider two alternative information assumptions, observable modes and unobservable modes. Under both alternatives, the central bank and the private sector know the probability distribution

[6] Without significant loss of generality, we could assume that the ε shocks are discrete, $\varepsilon_t \in \{\bar{\varepsilon}_h\}_{h=1}^{\bar{n}}$, and hence depend on separate modes $h = 1, ..., \bar{n}$ which may be correlated with the j modes. Then we could consider $n\bar{n}$ generalized modes (j, h) $(j = 1, ..., n,\ h = 1, ..., \bar{n})$ and incorporate the ε shocks in intercepts that depend on the generalized modes. This way we could, without loss of generality, write the model without any explicit additive ε shocks.

[7] When $\delta = 1$, the loss function (2.4) normally becomes unbounded. To handle this case, we scale the intertemporal loss function by $1 - \delta$ for $\delta < 1$ and consider the loss function to be the limit
$\lim_{\delta \to 1} (1-\delta) \mathrm{E}_t \sum_{\tau=0}^{\infty} \delta^\tau L(X_{t+\tau}, x_{t+\tau}, i_{t+\tau}, j_{t+\tau})$. See appendix D for details.

of the innovation ε_t, the transition matrix P, and the n_j different values each of the matrices in (2.1), (2.2), and (2.5) can take. That is, they have the same view of the nature of the model uncertainty. Furthermore, in the beginning of period t, before the central bank sets the instruments, i_t, under the assumption of *observable modes* the central bank's and private sector's information set includes the mode j_t and more generally the whole history of past the realizations: $\{X_t, j_t, \varepsilon_t, X_{t-1}, j_{t-1}, \varepsilon_{t-1}, x_{t-1}, i_{t-1}, \ldots\}$. Hence, the conditional expectations operator, E_t, refers to expectations conditional on that information.

Under the assumption of *unobservable modes*, the central bank and private sector cannot observe the modes.[8] Their subjective probability distribution over the modes in period t is denoted $p_t \equiv (p_{1t}, ..., p_{n_j t})'$. As noted in the introduction, we do not consider the case where the central bank and the private sector update their subjective distribution of modes based on observations of the economy. While this case is important and is examined in Svensson and Williams [42], the learning which it implies introduces nonlinearities which reduces the tractability of the MJLQ framework. Moreover, the case without learning on that we focus on here provides a useful starting point to analyze the effects of learning. In this paper, we assume that the subjective distribution simply evolves according to the exogenous transition probabilities. Then, conditional on p_t in period t, the distribution of the modes in the future period $t+\tau$ is given by

$$p_{t+\tau} = (P')^\tau p_t \qquad (\tau \geq 0). \tag{2.6}$$

It is worth noting what type of belief specification underlies the assumption that the central bank and private sector do not learn from their beliefs. In general, this requires the central bank and private sector to have subjective beliefs which are dynamically inconsistent or differ from the true data-generating process. A first possibility, *independent draws*, is that the central bank and private sector (incorrectly) view the future modes $j_{t+\tau}$ as being drawn independently each period $t+\tau$ from the exogenous distribution $p_{t+\tau}$ given by (2.6) in period t. In particular, if $p_t = \bar{p}$, they view the exogenous distribution as being the stationary distribution $\bar{p}$ associated with the transition matrix P. For this possibility, there is no (perceived) gain from updating the beliefs from observations of the economy. Hence not updating beliefs is optimal for this subjective probability distribution. In this case, in the beginning of period t, before the central bank sets the instruments, the common information set is the same as under observable modes, except that all parties do not observe j_t. Instead, they believe that modes have a probability distribution p_t, and that future modes will be drawn independently according to the distribution given by (2.6). The conditional expectations operator, E_t, then refers to expectations conditional on that information and beliefs. In an earlier version of the paper we focused on the assumption of independent draws.

A second possibility, *forgetting the past*, suggested to us by Alexei Onatski, is that the central bank and private sector in period t forget past observations of the economy, such as $\{X_{t-1}, X_{t-2}, \ldots\}$,

[8] Since forward-looking variables will be allowed to depend on the mode, *parts* of the private sector, but not the *aggregate* private sector, may be able to observe the mode.

when making decisions in period t. Without past observations, the policymaker cannot use current observations to update the beliefs. This possibility has the advantage that the policymaker need not view the modes as being independently drawn and may exploit the fact that the true modes may be serially correlated. However, forgetting past observations implies that the beliefs do not satisfy the law of iterated expectations, which requires the slightly complicated Bellman equations and derivations below. Under unobservable modes and forgetting the past, in the beginning of period t, before the central bank chooses the instruments, i_t, the common information set includes X_t and p_t, along with the fact that future modes will be distributed according to (2.6) and possibly be serially correlated depending on the matrix P. The conditional expectations operator then refers to expectations conditional on that information and those beliefs. We will employ the assumption of forgetting the past in our analysis below.

Under either observable or unobservable modes, we consider the optimization problem of minimizing (2.4) in period t, subject to (2.1), (2.2), (2.5), with either (X_t, j_t) or (X_t, p_t) given. We focus on optimization under commitment in a timeless perspective (see Woodford [44] and Svensson and Woodford [43]), although our results do not depend on this. As explained below, we will then add the term

$$\Xi_{t-1} \frac{1}{\delta} \mathrm{E}_t H_{j_t} x_t \tag{2.7}$$

to the intertemporal loss function in period t, where the elements of the n_x-vector Ξ_{t-1} are the Lagrange multipliers for equations (2.2) from the optimization problem in period $t-1$.

2.3 Reformulation according to the recursive saddlepoint method

As mentioned above, there has been extensive research in control theory developing methods for MJLQ systems. In order to apply those methods, we require that the system be recursive. However, the presence of the forward-looking variables in (2.2) makes the problem nonrecursive. Fortunately, the recursive saddlepoint method of Marcet and Marimon [29] can be applied to reformulate the non-recursive problem with forward-looking variables as a recursive saddlepoint problem (see Marcet and Marimon [29] for the general method and Svensson [40] for details of the method applied to linear-quadratic problems). This method adds lagged Lagrange multipliers corresponding to the forward-looking equations as additional state variables in the optimization problem. It also makes the current value of these multipliers, as well as the current values of the forward-looking variables themselves, additional control variables. By thus expanding the state and control space, we convert the non-recursive problem to a recursive one.

It will be practical to replace equation (2.2) by the two equivalent equations,

$$\mathrm{E}_t H_{j_{t+1}} x_{t+1} = z_t, \tag{2.8}$$

$$0 = A_{21 j_t} X_t + A_{22 j_t} x_t - z_t + B_{2 j_t} i_t, \tag{2.9}$$

where we introduce the n_x-vector of additional forward-looking variables, z_t. Introducing this vector is a practical way of keeping track of the expectations term on the left side of (2.2). Furthermore, it will be practical to use (2.9) and solve x_t as a function of X_t, z_t, i_t, and j_t,

$$x_t = \tilde{x}(X_t, z_t, i_t, j_t) \equiv A_{22j_t}^{-1}(-A_{21j_t}X_t + z_t - B_{2j_t}i_t). \tag{2.10}$$

We note that, for given j_t, this function is linear in X_t, z_t, and i_t.

For the application of the recursive saddlepoint method, the dual period loss function can be written

$$\mathrm{E}_t\tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j_t) \equiv \sum_j p_{jt}\tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j),$$

where $\tilde{X}_t \equiv (X_t', \Xi_{t-1}')'$ is the $n_{\tilde{X}}$-vector ($n_{\tilde{X}} \equiv n_X + n_x$) of extended predetermined variables (that is, including the n_x-vector Ξ_{t-1}), γ_t is an n_x-vector of Lagrange multipliers, and where

$$\tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j_t) \equiv L[X_t, \tilde{x}(X_t, z_t, i_t, j_t), i_t, j_t] - \gamma_t' z_t + \Xi_{t-1}' \frac{1}{\delta} H_{j_t}\tilde{x}(X_t, z_t, i_t, j_t). \tag{2.11}$$

For future reference, we can also write

$$\tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j_t) \equiv \begin{bmatrix} \tilde{X}_t \\ \tilde{\imath}_t \end{bmatrix}' \tilde{W}_{j_t} \begin{bmatrix} \tilde{X}_t \\ \tilde{\imath}_t \end{bmatrix}, \tag{2.12}$$

where $\tilde{\imath}_t \equiv (z_t', x_t', i_t')'$ and $\tilde{W}_{j_t}$ is an $(n_{\tilde{X}} + n_{\tilde{\imath}}) \times (n_{\tilde{X}} + n_{\tilde{\imath}})$ matrix, where $n_{\tilde{\imath}} \equiv n_x + n_x + n_i$. Furthermore, the case of unobservable modes corresponds to a given subjective probability distribution over modes, $p_t = (p_{1t}, ..., p_{n_jt})'$, whereas the case of observable modes corresponds to $p_{jt} = 1$ for $j = j_t$, $p_{jt} = 0$ for $j \neq j_t$ if $j_t \in N_j$ is the true mode. Then the dual intertemporal loss function is

$$\mathrm{E}_t \sum_{\tau=0}^{\infty} \delta^\tau \tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j_t), \tag{2.13}$$

and the dual optimization problem is find a saddlepoint such that (2.13) is maximized over $\{\gamma_{t+\tau}\}_{\tau\geq 0}$ and minimizes over $\{z_{t+\tau}, i_{t+\tau}\}_{\tau \geq 0}$ subject to the relevant transition equation.

2.3.1 Observable modes

Under observable modes, the state of the economy in period t is $(\tilde{X}_t, j_t)$, and the Bellman equation for the recursive saddlepoint problem (the dual optimization problem) with the value function $\tilde{V}(\tilde{X}_t, j_t)$ is

$$\begin{aligned} \tilde{V}(\tilde{X}_t, j_t) &= \max_{\gamma_t} \min_{(z_t, i_t)} [\tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j_t) + \mathrm{E}_t \delta \tilde{V}(\tilde{X}_{t+1}, j_{t+1})] \\ &\equiv \max_{\gamma_t} \min_{(z_t, i_t)} \left[\tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j_t) + \delta \sum\nolimits_{j_{t+1}} P_{j_t j_{t+1}} \int \tilde{V}(\tilde{X}_{t+1}, j_{t+1})\varphi(\varepsilon_{t+1}) d\varepsilon_{t+1}\right], \end{aligned} \tag{2.14}$$

where $\varphi(\cdot)$ denotes a generic probability density function. The optimization is subject to the transition equation for X_{t+1},

$$X_{t+1} = A_{11j_{t+1}} X_t + A_{12j_{t+1}} \tilde{x}(X_t, z_t, i_t, j_t) + B_{1j_{t+1}} i_t + C_{j_{t+1}} \varepsilon_{t+1}, \tag{2.15}$$

where we have substituted $\tilde{x}(X_t, z_t, i_t, j_t)$ for x_t, and the new dual transition equation for Ξ_t,

$$\Xi_t = \gamma_t. \tag{2.16}$$

With the new states and controls, we then have a relatively standard optimization problem. If the modes were fixed, this would simply be a standard LQ problem whose well-known solution is a quadratic value function and a linear policy function. However the switching of the modes adds an important nonlinearity. Nonetheless, the evolution is linear and preferences are quadratic conditional on the modes, and the evolution of the modes is independent of the predetermined variables X_t. This leads to the value function being quadratic and the optimal policy linear conditional on the modes. In particular, as is shown in appendix B, the dual value function $\tilde{V}(\tilde{X}_t, j_t)$ is quadratic in $\tilde{X}_t$ for given j_t, taking the form

$$\tilde{V}(\tilde{X}_t, j_t) \equiv \tilde{X}_t' \tilde{V}_{\tilde{X}\tilde{X}j_t} \tilde{X}_t + w_{j_t}, \quad j_t \in N_j, \tag{2.17}$$

where $\tilde{V}_{\tilde{X}\tilde{X}j_t}$ is an $n_{\tilde{X}} \times n_{\tilde{X}}$ matrix and w_{j_t} is a scalar. In addition, the optimal policies are linear in $\tilde{X}_t$ for given $j_t \in N_j$,

$$\tilde{\imath}_t \equiv \begin{bmatrix} z_t \\ i_t \\ \gamma_t \end{bmatrix} = \tilde{\imath}(\tilde{X}_t, j_t) \equiv \begin{bmatrix} z(\tilde{X}_t, j_t) \\ i(\tilde{X}_t, j_t) \\ \gamma(\tilde{X}_t, j_t) \end{bmatrix} = F_{j_t} \tilde{X}_t \equiv \begin{bmatrix} F_{zj_t} \\ F_{ij_t} \\ F_{\gamma j_t} \end{bmatrix} \tilde{X}_t, \tag{2.18}$$

$$x_t = x(\tilde{X}_t, j_t) \equiv \tilde{x}(X_t, z(\tilde{X}_t, j_t), i(\tilde{X}_t, j_t), j_t) \equiv F_{xj_t} \tilde{X}_t. \tag{2.19}$$

This solution is also the solution to the primal optimization problem. The equilibrium transition equation is then given by

$$\begin{aligned} \tilde{X}_{t+1} &\equiv \begin{bmatrix} X_{t+1} \\ \Xi_t \end{bmatrix} = \begin{bmatrix} A_{11j_{t+1}} X_t + A_{12j_{t+1}} F_{xj_t} \tilde{X}_t + B_{1j_{t+1}} F_{ij_t} \tilde{X}_t \\ F_\gamma(j_t) \tilde{X}_t \end{bmatrix} + \begin{bmatrix} C_{1j_{t+1}} \\ 0 \end{bmatrix} \varepsilon_{t+1} \\ &\equiv M_{j_t j_{t+1}} \tilde{X}_t + \tilde{C}_{j_{t+1}} \varepsilon_{t+1}, \end{aligned} \tag{2.20}$$

where $M_{j_t j_{t+1}}$ is an $n_{\tilde{X}} \times n_{\tilde{X}}$ matrix and $\tilde{C}_{j_{t+1}}$ is an $n_{\tilde{X}} \times n_X$ matrix. Details and a convenient algorithm for computing $\tilde{V}_{\tilde{X}\tilde{X}j}$ and F_j for $j \in N_j$ are provided in appendix B.

Consider the composite state $(\tilde{X}_t, j_t)$ in period t, where $\tilde{\imath}_t = F_{j_t} \tilde{X}_t$. The transition from this composite state to the composite state $(\tilde{X}_{t+1}, j_{t+1})$ in period $t+1$ with $\tilde{\imath}_{t+1} = F_{j_{t+1}} \tilde{X}_{t+1}$ will satisfy (2.20) and will, for given realization of ε_{t+1}, occur with probability $P_{j_t j_{t+1}}$. This determines the optimal distribution of future $\tilde{X}_{t+\tau}$, $j_{t+\tau}$, and $\tilde{\imath}_{t+\tau}$ $(\tau \geq 1)$ conditional on $(\tilde{X}_t, j_t)$. Such conditional distributions can be illustrated by plots of future means, medians, and percentiles (fan charts).

Plots of future means, medians, and percentiles can also be constructed for individual chains of the modes, for instance, the median or mean chain corresponding to no model uncertainty. The simplest way to generate such plots is by simulation, which we illustrate in some examples below.

Note that the value function in (2.17) above corresponds to the dual period loss function and the dual saddlepoint problem. The primal value function $V(\tilde{X}_t, j_t)$ for the original problem of minimizing (2.4) subject to (2.1), (2.2), and (2.5) under commitment in a timeless perspective satisfies

$$V(\tilde{X}_t, j_t) \equiv \tilde{X}_t' V_{\tilde{X}\tilde{X}j_t} \tilde{X}_t + w_{j_t} \equiv V(\tilde{X}_t, j_t) - \frac{1}{\delta}\Xi_{t-1}' H_{j_t} F_{xj_t} \tilde{X}_t \equiv \tilde{X}_t' \tilde{V}_{\tilde{X}\tilde{X}j} \tilde{X}_t + w_{j_t} - \frac{1}{\delta}\Xi_{t-1}' H_{j_t} F_{xj_t} \tilde{X}_t, \tag{2.21}$$

where we note that the scalar w_{j_t} is the same for the dual and primal value functions. Since

$$\frac{1}{\delta}\Xi_{t-1}' H_j F_{xj} \tilde{X}_t \equiv \frac{1}{2\delta}(\Xi_{t-1}' H_j F_{xj} \tilde{X}_t + \tilde{X}_t' F_{xj}' H_j' \Xi_{t-1}) \equiv \tilde{X}_t' \begin{bmatrix} 0 & \frac{1}{2\delta} F_{xXj}' H_j' \\ \frac{1}{2\delta} H_j F_{xXj} & \frac{1}{2\delta}(H_j F_{x\Xi j} + F_{x\Xi j}' H_j') \end{bmatrix} \tilde{X}_t,$$

where F_{xj} is partitioned conformably with X_t and Ξ_{t-1} as $F_{xj} \equiv [F_{xXj} \;\; F_{x\Xi j}]$, the matrices $V_{\tilde{X}\tilde{X}j}$ are given by

$$V_{\tilde{X}\tilde{X}j} = \tilde{V}_{\tilde{X}\tilde{X}j} - \begin{bmatrix} 0 & \frac{1}{2\delta} F_{xXj}' H_j' \\ \frac{1}{2\delta} H_j F_{xXj} & \frac{1}{2\delta}(H_j F_{x\Xi j} + F_{x\Xi j}' H_j') \end{bmatrix} \qquad (j \in N_j). \tag{2.22}$$

As discussed in CFM [11], mean square stability is an appropriate concept of stability for the observable-modes case. Appendix E provides some details on the definition of mean square stability and shows how the necessary and sufficient conditions for mean square stability derived in CFM [11] can be applied in our case.

2.3.2 Unobservable modes

Under unobservable modes and forgetting the past, let $s_t \equiv (\tilde{X}_t', p_t')'$ denotes the *perceived* state of the economy ("perceived" in the sense that it includes the perceived probability distribution, p_t, but not the true mode). We find it useful to introduce the conditional dual value function $\hat{V}(s_t, j_t)$, which gives the dual intertemporal loss conditional on the true state of the economy, (s_t, j_t). This function satisfies

$$\hat{V}(s_t, j) \equiv \int \begin{bmatrix} \tilde{L}(\tilde{X}_t, z(s_t), i(s_t), \gamma(s_t), j) \\ +\,\delta \sum_k P_{jk} \hat{V}[\bar{g}(s_t, j, \varepsilon_t, k, \varepsilon_{t+1}), k] \end{bmatrix} \varphi(\varepsilon_{t+1}) d\varepsilon_{t+1} \qquad (j \in N_j).$$

The true dual value function $\tilde{V}(s_t)$ for the problem averages over the conditional value functions according to the perceived distribution of modes:

$$\tilde{V}(s_t) = \mathrm{E}_t \hat{V}(s_t, j_t) = \sum_j p_{jt} \hat{V}(s_t, j).$$

Then $\tilde{V}(s_t)$ solves the somewhat unusual Bellman equation:

$$\begin{aligned}\tilde{V}(s_t) &= \max_{\gamma_t} \min_{(z_t,i_t)} \mathrm{E}_t\{\tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j_t) + \delta\hat{V}[g(s_t, z_t, i_t, \gamma_t, j_t, j_{t+1}, \varepsilon_{t+1}), j_{t+1}]\} \\ &\equiv \max_{\gamma_t} \min_{(z_t,i_t)} \sum\nolimits_j p_{jt} \int \left[\begin{array}{l} \tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j) \\ +\,\delta \sum_k P_{jk}\hat{V}[g(s_t, z_t, i_t, \gamma_t, j, k, \varepsilon_{t+1}), k] \end{array} \right] \varphi(\varepsilon_{t+1}) d\varepsilon_{t+1}, \end{aligned} \tag{2.23}$$

where not only the perceived state of the economy, s_t, appears, but also the *true* state of the economy, (s_t, j_t) ("true" in the sense that it includes the true mode of the economy). For the case of unobservable modes and forgetting the past, it is necessary to include the mode j_t in the state vector because the beliefs do not satisfy the law of iterated expectations. In the Bellman equation we require that the solution for z_t, i_t, and γ_t respect the information constraints and thus depend on the perceived state s_t but not directly on the mode j_t.

The optimization is subject to (2.15), (2.16), and the transition equation for p_{t+1}, (2.6). This can be combined into the transition equation for s_{t+1},

$$\begin{aligned} s_{t+1} \equiv \begin{bmatrix} X_{t+1} \\ \Xi_t \\ p_{t+1} \end{bmatrix} &= g(s_t, z_t, i_t, \gamma_t, j_t, j_{t+1}, \varepsilon_{t+1}) \\ &\equiv \begin{bmatrix} A_{11j_{t+1}}X_t + A_{12j_{t+1}}\tilde{x}(X_t, z_t, i_t, j_t) + B_{1j_{t+1}}i_t + C_{1j_{t+1}}\varepsilon_{t+1} \\ \gamma_t \\ P'p_t \end{bmatrix}. \end{aligned} \tag{2.24}$$

In the case of observable modes above, the solution to the dual problem was a mode-dependent quadratic value function and mode-dependent linear policy. These results relied in an important way on the evolution of the modes being exogenous. Under unobservable modes but forgetting the past, the belief evolution is also exogenous, and thus a similar argument applies and we obtain a belief-dependent quadratic value function and a belief-dependent linear policy. In particular, it is straightforward to see that the solution of the dual optimization problem is linear in $\tilde{X}_t$ for given s_t,

$$\tilde{\imath}_t \equiv \begin{bmatrix} z_t \\ i_t \\ \gamma_t \end{bmatrix} = \tilde{\imath}(s_t) \equiv \begin{bmatrix} z(s_t) \\ i(s_t) \\ \gamma(s_t) \end{bmatrix} = F(p_t)\tilde{X}_t \equiv \begin{bmatrix} F_z(p_t) \\ F_i(p_t) \\ F_\gamma(p_t) \end{bmatrix} \tilde{X}_t, \tag{2.25}$$

$$x_t = x(s_t, j_t) \equiv \tilde{x}(X_t, z(s_t), i(s_t), j_t) \equiv F_x(p_t, j_t)\tilde{X}_t, \tag{2.26}$$

where the forward-looking variables also depend on the mode j_t.[9] This solution is also the solution to the primal optimization problem. The equilibrium transition equation is then given by

$$s_{t+1} = \bar{g}(s_t, j_t, j_{t+1}, \varepsilon_{t+1}) \equiv g[s_t, z(s_t), i(s_t), \gamma(s_t), j_t, j_{t+1}, \varepsilon_{t+1}].$$

This equilibrium transition equation together with (2.25) and (2.26) can be used to construct conditional distributions of future states of the economy. The simplest way to generate such plots is by simulation, which we illustrate in some examples below.

[9] We assume that the central bank and private sector do not update the probability distribution over modes from observation of the forward-looking variables. Updating the probability distribution from observations of X_t and x_t is examined in Svensson and Williams [42].

Moreover, the (unconditional) dual value function $\tilde{V}(s_t)$ is quadratic in $\tilde{X}_t$ for given p_t, taking the form

$$\tilde{V}(s_t) \equiv \tilde{X}_t' \tilde{V}_{\tilde{X}\tilde{X}}(p_t)\tilde{X}_t + w(p_t).$$

The function $\hat{V}(s_t, j_t)$ is also quadratic in $\tilde{X}_t$ for given p_t and j_t,

$$\hat{V}(s_t, j_t) \equiv \tilde{X}_t' \hat{V}_{\tilde{X}\tilde{X}}(p_t, j_t)\tilde{X}_t + \hat{w}(p_t, j_t).$$

It follows that we have

$$\tilde{V}_{\tilde{X}\tilde{X}}(p_t) \equiv \sum\nolimits_j p_{jt}\hat{V}_{\tilde{X}\tilde{X}}(p_t, j), \qquad w(p_t) \equiv \sum\nolimits_j p_{jt}\hat{w}(p_t, j).$$

The value function for the primal problem, with the period loss function $\mathrm{E}_t L(X_t, x_t, i_t, j_t)$ rather than $\mathrm{E}_t \tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j_t)$, satisfies

$$V(s_t) \equiv \tilde{V}(s_t) - \Xi_{t-1}' \frac{1}{\delta} \sum_j p_{jt} H_j x(s_t, j). \tag{2.27}$$

It is again quadratic in $\tilde{X}_t$ for given p_t,

$$V(s_t) \equiv \tilde{X}_t' V_{\tilde{X}\tilde{X}}(p_t)\tilde{X}_t + w(p_t)$$

(the scalar $w(p_t)$ in the primal value function is obviously identical to that in the dual value function). This is the value function conditional on $\tilde{X}_t$ and p_t, taking into account that j_t is not observable. Hence, the second term on the right side of (2.27) contains the expectation of $H_{j_t} x_t$ conditional on that information. The matrix $V_{\tilde{X}\tilde{X}}(p_t)$ is related to the matrix $\tilde{V}_{\tilde{X}\tilde{X}}(p_t)$ in a way comparable to (2.22).

Algorithms for determining the solution and the value functions are presented in appendix C. Computing the functions $F(p_t)$, $V_{\tilde{X}\tilde{X}}(p_t)$, and $F(p_t)$ for all feasible values of p_t requires standard function-approximation methods, as they are vector- and matrix-valued functions of the probabilities. However, as shown in appendix C, computing the functions for a particular value p_t is straightforward.[10]

3 Interpretation of model uncertainty in our framework

As we've seen, the assumption that the random matrices of coefficients take a finite number of values corresponding to a finite number of modes allows us to use the convenient and flexible

[10] Consider the degenerate distributions, $p_t = e_j$ where e_j is the distribution where $p_j = 1$, $p_k = 0$ ($k \neq j$). That is, $p_t = e_j$ corresponds to the case when the mode j is observed in period t. Note that $V_{\tilde{X}\tilde{X}}(e_j) \neq V_{\tilde{X}\tilde{X}j}$ and $F(e_j) \neq F_j$, where $V_{\tilde{X}\tilde{X}j}$ and F_j ($j = 1, ..., n$) denote the value function and optimal policy function matrices for observable modes, when the modes are observed in *each* period. The reason is that even if $p_t = e_j$ and the mode is observed in period t, the distribution of the modes in the next period $t+1$ will be $p_{t+1} = e_j P = (P_{j1}, P_{j2}, ..., P_{jn})$ and the modes will not be observed in the next period. In contrast, $V_{\tilde{X}\tilde{X}j}$ and F_j are derived under the assumption that the modes are observed in all periods.

framework of MJLQ systems. In order to do so, we apply the recursive saddlepoint method of Marcet and Marimon to reformulate the non-recursive model with forward-looking variables as a recursive model. The flexibility of the MJLQ approach allows us to consider a wide variety of applications. By specifying different configurations of modes and transition probabilities, we can approximate many different kinds of model uncertainty.

- Both i.i.d. and serially correlated random coefficients of the model can be handled. This can capture either generalized parameter uncertainty or different behavior in different modeled regimes (such boom/bust states, and so forth). In addition, switches in the underlying processes of nonlinear models (such as in the mean growth rate of productivity) can be linearized to yield MJLQ systems.

- The modes can correspond to different structural models. The models can differ by having different relevant variables, different number of leads or lags, or the same variable being predetermined in one model and forward-looking in another. For example, one mode can represent a model with forward-looking variables such as the New Keynesian model of Lindé [27], another a backward-looking model such as that of Rudebusch and Svensson [33]. We consider such an example below.

- The modes can correspond to situations when variables such as inflation and output have more or less inherent persistence (are more or less autocorrelated), when the exogenous shocks have more or less persistence (introduce a predetermined variable equal to the serially correlated shock, letting it be an AR(1) process with a high or low coefficient), or when the uncertainty about the coefficients or models is higher or lower.

- The modes can be structured such that they correspond to different central-bank judgments about model coefficients and model uncertainty. Let $j_t = 1, ..., n_j$ correspond to n_j different *model* modes (different coefficients, different variances or persistence of coefficient disturbances, or different variances of the ε_t shocks). Let $k_t = 1, ..., n_k$ correspond to n_k different central-bank *judgment* modes, depicting some central-bank information about the model modes. This can generally be modeled as a situation where the transition matrix for the model modes depends on the judgment mode. Thus let the transition matrix for model modes be $\tilde{P}(k_t)$, and hence depend on k_t. Let P^0 denote the transition matrix for the judgment modes (assumed independent of the model modes). We can then consider a composite model-judgment mode (j_t, k_t) in period t, with the transition probability from model-judgment mode (h, k) in period t to mode (j, l) in period $t+1$ given by $\tilde{P}(k)_{hj} P^0_{kl}$. For instance, the judgment modes may correspond to different persistence of the model modes.

- As noted in appendix A, we can combine multiplicative uncertainty about the modes with the additive uncertainty about future deviations. This way we can simultaneously handle central-bank judgment about future additive deviations as in Svensson [39] and central-bank judgment about model modes as in this paper. For instance, we can handle situations when there is more or less uncertainty about shocks farther into the future relative to those in the near future.

Generally, aside from dimensional and computational limitations, it is difficult to conceive of a relevant situation for a policymaker that cannot be approximated in this framework. Moreover, the computational constraints are not overly tight. The examples below have up to 9 state variables and 3 modes and are solvable in seconds. We have also considered examples with up to 50 modes and 10 state variables which were solvable in a few minutes. For a one-time policy optimization, such speed and flexibility is useful, but for in-depth policy analysis and estimation (considering different scenarios, model specifications, and loss functions) it is critical.

4 Examples

In this section we present examples based on two empirical models of the US economy: regime-switching versions of the backward-looking model of Rudebusch and Svensson [33] and the forward-looking New Keynesian model of Lindé [27]. These examples can be interpreted as model uncertainty within a given class of structural models, in this case a setting with exogenous switches in the underlying parameters of the models. We then turn to an example which has more of a structural model-uncertainty flavor, where we re-estimate the Lindé [27] model constraining one mode to be backward-looking while the other has both backward- and forward-looking elements. This corresponds to a situation when there is uncertainty about the importance of forward-looking behavior.

4.1 An estimated backward-looking model

In this section we consider the effects of model uncertainty in the quarterly model of the US economy of Rudebusch and Svensson [33], henceforth RS. The key variables in the model are quarterly annualized inflation π_t, the output gap y_t, and the annualized instrument rate i_t. We use data obtained from the St. Louis Fed FRED website covering 1961:Q1-2006:Q1, taking the chain-weighted GDP deflator as our price index, the percentage deviation of GDP from the CBO estimate of potential as our measure of the output gap, and the federal funds rate as the instrument rate. This same data set is used in our other examples below. In this section, we estimate a three-mode MJLQ model using Bayesian methods to locate the maximum of the posterior distribution, and we compare the implications to the constant-coefficient version of RS.

Parameter	Constant	Mode 1	Mode 2	Mode 3
α_0	0.5697	0.3744	0.6598	0.5437
α_1	0.0752	0.1336	0.0329	0.0678
α_2	0.1276	0.1524	0.1362	0.0999
α_3	0.1451	0.1099	0.1652	0.1029
β_1	1.1834	1.2417	1.1551	1.2162
β_2	−0.2651	−0.3408	−0.2398	−0.2717
β_3	−0.0510	−0.0115	−0.0393	−0.0206
c_π	1.0070	0.7276	1.4008	0.6936
c_y	0.7540	0.4748	1.0777	0.7445

Table 4.1: Estimates of the constant-coefficient and three-mode Rudebusch-Svensson model.

The model has a Phillips curve and an aggregate-demand relation of the following form:

$$\begin{aligned} \pi_{t+1} &= \sum_{\tau=0}^{2} \alpha_{\tau j}\pi_{t-\tau} + (1 - \sum_{\tau=0}^{2} \alpha_{\tau j})\pi_{t-3} + \alpha_{3j} y_t + c_{\pi j}\varepsilon_{\pi,t+1}, \\ y_{t+1} &= \beta_{1j} y_t + \beta_{2j} y_{t-1} + \beta_{3j} \left(\bar{\imath}_t - \bar{\pi}_t\right) + c_{yj}\varepsilon_{y,t+1}, \end{aligned} \tag{4.1}$$

where $j \in \{1, 2, 3\}$ indexes the mode, $\bar{\imath}_t \equiv \sum_{\tau=0}^{3} i_{t-\tau}/4$ and $\bar{\pi}_t \equiv \sum_{\tau=0}^{3} \pi_{t-\tau}/4$ are 4-quarter averages, and the shocks $\varepsilon_{\pi t}$ and ε_{yt} are each independent standard normal random variables.

Table 4.1 reports our estimates of the peak of the posterior, with the OLS estimates of the constant-coefficient version of the model for comparison. For the MJLQ model, we center our prior distribution at the OLS estimates and restrict α_3 to be positive and β_3 to be negative. Details of the estimation method and prior setting are given in appendix F. Here we see that many of the coefficients differ substantially across modes. Perhaps most notable is the large difference in volatility, as the standard deviation of the inflation shocks (c_π) in mode 2 is more than twice as large as the estimates in the other modes, and the output gap shocks (c_y) are largest in this mode as well. In addition, the slope of the aggregate-demand relation, β_3, ranges from a relatively large negative response in mode 2 to a small negative one in mode 1. The differences in these key model coefficients lead to some differences in the optimal policy across modes for the observable-modes case, as we show below.

The estimated probabilities of being in the different modes are shown in figure 4.1. The plots show both the filtered estimates, in which the distribution in period t is estimated using data only up to t, as well as the smoothed estimates, in which the distribution in period t is estimated using data for the whole sample. Clearly, there are more fluctuations in the filtered estimates than in the smoothed ones, since by looking backward we can better assess the probability of being in a particular regime. We see that, for the very early part of the sample, the economy was mostly assessed to be in the relatively calm mode 3. But then throughout the 1970s until the early 1980s the economy was mostly in the more volatile mode 2. From the early 1980s onward, mode 1 has been predominant, as the volatility moderated. The estimated transition matrix P and its implied

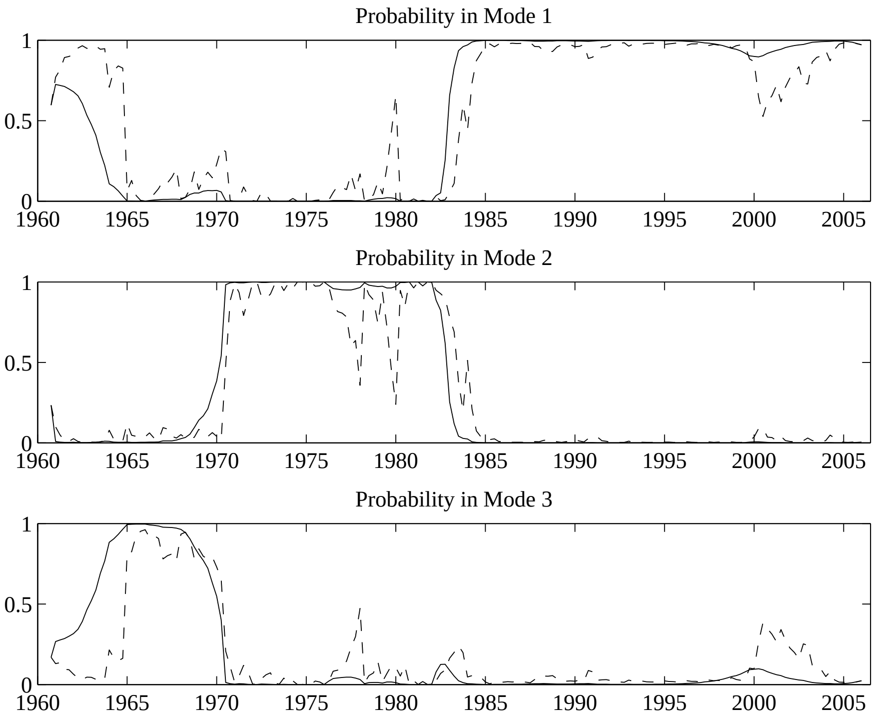

Figure 4.1: Estimated probabilities of the different modes. Solid lines: Smoothed (full-sample) inference. Dashed lines: Filtered (one-sided) inference.

stationary distribution $\bar{p}$ are

$$P = \begin{bmatrix} 0.9887 & 0.0056 & 0.0057 \\ 0.0145 & 0.9711 & 0.0143 \\ 0.0199 & 0.0201 & 0.9601 \end{bmatrix}, \qquad \bar{p} = \begin{bmatrix} 0.5967 \\ 0.2340 \\ 0.1694 \end{bmatrix}.$$

From the standpoint of these estimates, the data from the 1980s on is the most typical, as mode 1 has the highest weight in the stationary distribution. Similar episodes will thus re-occur in the model, although balanced with periods of larger volatility.

We let the period loss function be

$$L_t = \pi_t^2 + \lambda y_t^2 + \nu(i_t - i_{t-1})^2. \tag{4.2}$$

Hence, the vector of target variables is $Y_t \equiv (\pi_t, y_t, i_t - i_{t-1})'$ and the weight matrix Λ is a diagonal matrix with the diagonal $(1, \lambda, \nu)'$. We set the weights to $\lambda = 1$ and $\nu = 0.2$, and fix set the discount factor in the intertemporal loss function to $\delta = 1$. Using the methods described above, we then solve for the optimal policy functions for the case of observable modes and the case of unobservable modes,

$$\begin{aligned} i_t &= F_{ij} X_t, \qquad (j = 1, 2, 3), \\ i_t &= F_i(\bar{p}) X_t, \end{aligned}$$

Mode	π_t	π_{t-1}	π_{t-2}	π_{t-3}	y_t	y_{t-1}	i_{t-1}	i_{t-2}	i_{t-3}
Constant	1.1053	0.5037	0.4160	0.2665	2.1640	−0.5772	0.5120	−0.0549	−0.0278
Mode 1	0.8721	0.5456	0.4308	0.2976	1.6220	−0.5838	0.7821	−0.0106	−0.0051
Mode 2	1.3269	0.4851	0.4333	0.2440	2.4116	−0.5764	0.5625	−0.0456	−0.0232
Mode 3	1.0219	0.4750	0.4037	0.2982	2.2605	−0.6209	0.6786	−0.0240	−0.0119
Unobservable	0.9907	0.5289	0.4321	0.2920	1.9642	−0.6163	0.7001	−0.0209	−0.0102

Table 4.2: Optimal policy functions for the constant-coefficient and three-mode Rudebusch-Svensson model.

respectively, where $X_t \equiv (\pi_t, \pi_{t-1}, \pi_{t-2}, \pi_{t-3}, y_t, y_{t-1}, i_{t-1}, i_{t-2}, i_{t-3})'$. For the case of unobservable modes, we hence compute the optimal policy function for the stationary distribution, $\bar{p}$.

The optimal policy functions are given in table 4.2. The first row gives the optimal policy function for the constant-coefficient case, the next three rows give it for the observable-modes case, and the last row gives it for the unobservable-modes case. However the coefficients of the policy functions are in themselves a bit difficult to interpret. Thus we plot the distribution of the impulse responses of inflation, the output gap, and the instrument rate to the two shocks in the model in figure 4.2 for the observable-modes case and in figure 4.3 for the unobservable-modes case. In particular, for each of 10,000 simulation runs, we first draw an initial mode of the Markov chain from its stationary distribution, then simulate the chain for 50 periods forward, tracing out the impulse responses. The figure plots the mean and median responses at each date, along with 30% quantiles of the empirical distribution. More precisely, the dark, medium, and light grey band show 30%, 60%, and 90% probability bands, respectively, with 5% of the distribution above the light gray band and 5% below. Also shown for comparison are the responses under the optimal policy for the estimated constant-coefficient model given above.[11]

The impulse responses for observable and unobservable modes are in this case quite similar. Both the table and the figures illustrate that the model uncertainty leads to a change in the nature of policy. Compared to the constant-coefficient model, most of the mass of the distribution of the impulse responses for the first few quarters lies closer to zero. This is particularly the case for the instrument-rate responses. Thus our results here are in accord with the common intuition based on Brainard [6], that model uncertainty should lead to less aggressive (that is, smaller in magnitude) policy responses.[12] Interestingly, the probability distributions of responses are asymmetric, with the mean impulse responses quite different from the median responses. These results illustrate that with model uncertainty policy makers must go beyond forecasting the means of target variables and consider the entire forecast distributions. Our approach makes this process quite manageable.

[11] The shocks are $\varepsilon_{\pi 0} = 1$ and $\varepsilon_{y0} = 1$, respectively, for the two different columns in the figure. Thus the effective shocks to inflation and the output gap in period 0 are mode dependent and equal to $c_{\pi j}$ and c_{yj} $(j = 1, 2, 3)$, respectively. As we initialize by drawing from the stationary distribution, the distribution of modes in each period remains the stationary distribution.

[12] Of course, this is only a loose parallel, as the Brainard result need not apply for the type of uncertainty considered here, especially in the observable-modes case when the policy is allowed to be mode-dependent. In addition, the means of the MJLQ coefficients do not equal the constant coefficients.

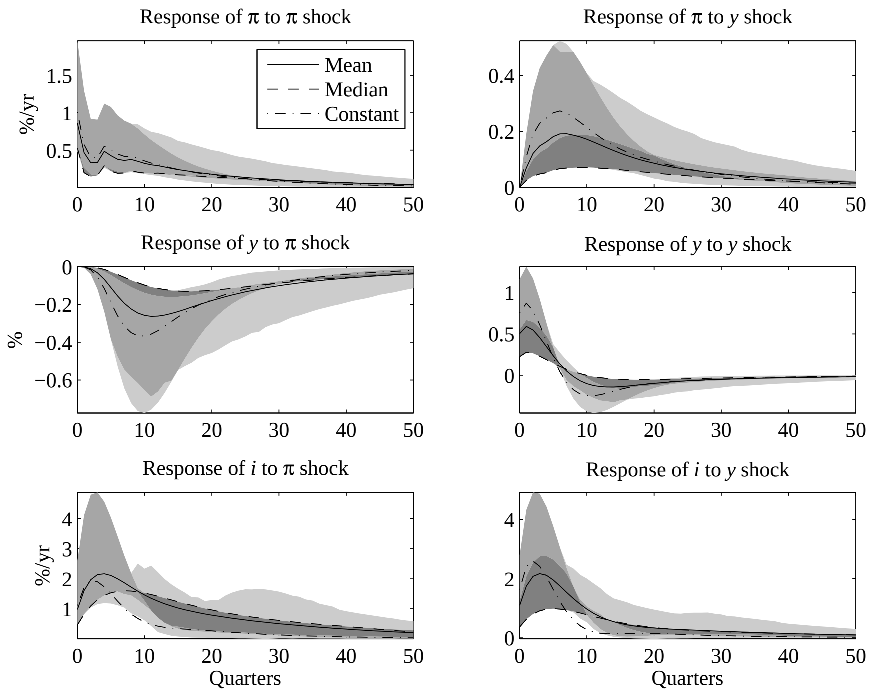

Figure 4.2: Unconditional impulse responses to shocks under the optimal policy for the mode-dependent Rudebusch-Svensson model. Solid lines: Mean responses. Dashed lines: Median response. Dark/medium/light gray bands: 30/60/90% probability bands. Dashed-dotted lines: Optimal responses for the constant-coefficient model.

4.2 An estimated forward-looking model

We now consider the effects of uncertainty in a model with both forward- and backward-looking variables. The structural model is a mode-dependent simplification of the model of the US economy of Lindé [27] and is given by

$$\begin{aligned}
\pi_t &= \omega_{fj} E_t \pi_{t+1} + (1-\omega_{fj})\pi_{t-1} + \gamma_j y_t + c_{\pi j}\varepsilon_{\pi t}, \\
y_t &= \beta_{fj} E_t y_{t+1} + (1-\beta_{fj})\left[\beta_{yj} y_{t-1} + (1-\beta_{yj}) y_{t-2}\right] - \beta_{rj}\left(i_t - E_t \pi_{t+1}\right) + c_{yj}\varepsilon_{yt}, \\
i_t &= \left(1-\rho_{1j}-\rho_{2j}\right)\left(\gamma_{\pi j}\pi_t + \gamma_{yj} y_t\right) + \rho_{1j} i_{t-1} + \rho_{2j} i_{t-2} + c_{ij}\varepsilon_{it},
\end{aligned} \tag{4.3}$$

where an instrument rule is added to the Phillips curve and the aggregate-demand relation.[13] Again, $j \in \{1,2,3\}$ indexes the mode, and the shocks $\varepsilon_{\pi t}$, ε_{yt}, and ε_{it} are independent standard normal random variables. We use the same data set as above, and again estimate a three-mode MJLQ model along with a constant-coefficient model using Bayesian methods. Once again, we explicitly state our prior settings in appendix F. We use the same prior for the structural coefficients in the

[13] Because of the forward-looking expectations in the model, estimation of the model requires that a policy rule be specified.

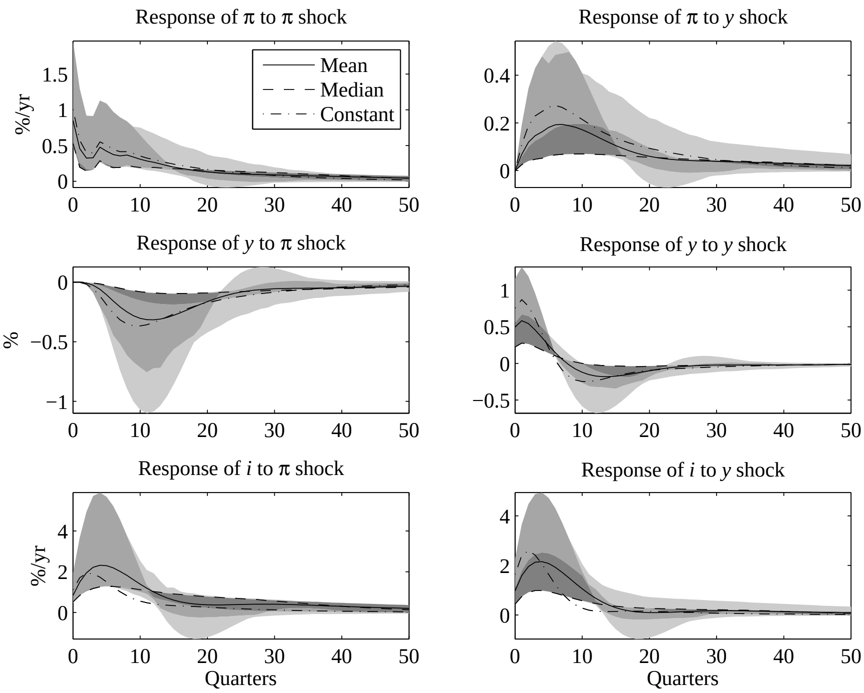

Figure 4.3: Unconditional impulse responses to shocks under the optimal policy for the unobservable-mode version of the Rudebusch-Svensson model. Solid lines: Mean responses. Dashed lines: Median responses. Dark/medium/light grey bands: 30/60/90% probability bands. Dashed-dotted lines: Optimal responses for constant coefficients.

constant-coefficient and MJLQ cases, and the priors for the Markov chain coefficients are the same as in the RS model.

Table 4.3 reports our estimates, with the estimates from the constant-coefficient version of the model for comparison. Our constant-coefficient estimates are similar to those in Lindé [27], with the main difference that we find much smaller estimates of γ and β_r. At least some of the difference may be due to our different data series and sample period. We again see that many of the key structural coefficients change substantially across modes, particularly the shock standard deviations. For example, mode 1 has the largest shocks to inflation and interest rates, while mode 2 has the smallest shocks to these variables. The degree of forward-looking behavior also varies across modes, with mode 2 having the lowest weights on the forward-looking terms in both key model equations. The slope of the Phillips curve γ and the interest sensitivity of demand β_r, which are key parameters governing how changes in the instrument rate are transmitted to inflation, also vary markedly across modes.

Parameter	Constant	Mode 1	Mode 2	Mode 3
ω_f	0.5164	0.3000	0.1496	0.5595
γ	0.0034	0.0643	0.0321	0.0205
β_f	0.4484	0.4595	0.0757	0.4139
β_r	0.0073	0.0067	0.0278	0.0902
β_y	1.1902	1.2943	1.2191	0.9310
ρ_1	0.8216	0.7051	1.3038	0.9994
ρ_2	0.0560	0.1349	−0.3748	−0.1744
γ_π	1.5538	1.6441	1.2982	0.7725
γ_y	0.9777	0.6967	0.6431	1.0784
c_π	0.5920	1.0378	0.6943	0.8076
c_y	0.3753	0.4763	0.5147	0.5740
c_i	1.0412	2.1133	0.3712	0.7073

Table 4.3: Estimates of the constant-coefficient and three-mode Lindé model.

Mode	π_{t-1}	y_{t-1}	y_{t-2}	i_{t-1}	$\varepsilon_{\pi t}$	ε_{yt}	$\Xi_{\pi,t-1}$	$\Xi_{y,t-1}$
Constant	0.1738	0.9394	−0.2112	0.7623	0.2128	0.7559	0.0011	0.0252
Mode 1	0.9582	0.9171	−0.3461	0.7446	1.4205	1.0362	0.0026	0.0215
Mode 2	1.9812	2.6539	−0.5794	0.6252	1.6176	1.4724	0.0004	0.0039
Mode 3	0.3556	0.8943	0.0639	0.4947	0.6520	0.9071	0.0066	0.0758
Unobservable	1.7877	2.2912	−0.4708	0.5968	1.6633	1.4033	0.0010	0.0142

Table 4.4: Optimal policy functions of the constant-coefficient and three-mode Lindé model.

The estimated transition matrix P and its implied stationary distribution $\bar{p}$ are given by

$$P = \begin{bmatrix} 0.9411 & 0.0294 & 0.0294 \\ 0.0053 & 0.9893 & 0.0054 \\ 0.0271 & 0.0262 & 0.9468 \end{bmatrix}, \qquad \bar{p} = \begin{bmatrix} 0.1322 \\ 0.7214 \\ 0.1464 \end{bmatrix}.$$

Thus mode 2 is the most persistent and has the largest mass in the invariant distribution. This is consistent with our estimation of the modes as shown in figure 4.4. Again, the plots show both the smoothed and filtered estimates. We see that mode 2 was experienced the most throughout much of the sample, holding for 1961-1968 and then with near certainty continually since 1985. The volatile mode 1 held for much of the early 1970s and early 1980s, alternating with the intermediate mode 3.

Using the methods described above, we again solve for the optimal policy functions,

$$\begin{aligned} i_t &= F_{ij}\tilde{X}_t, \\ i_t &= F_i(\bar{p})\tilde{X}_t, \end{aligned}$$

for the cases of observable and unobservable modes, respectively, where now $\tilde{X}_t \equiv (\pi_{t-1}, y_{t-1}, y_{t-2}, i_{t-1}, \varepsilon_{\pi t}, \varepsilon_{yt}, \Xi_{\pi,t-1}, \Xi_{y,t-1})'$. We use the same loss function as in the backward-looking model. The optimal policy functions are given in table 4.4. For ease of interpretation, as above we plot the distribution of the impulse responses of inflation, the output gap, and the instrument rate to the two structural shocks in the model for observable modes in figure 4.5 and unobservable modes

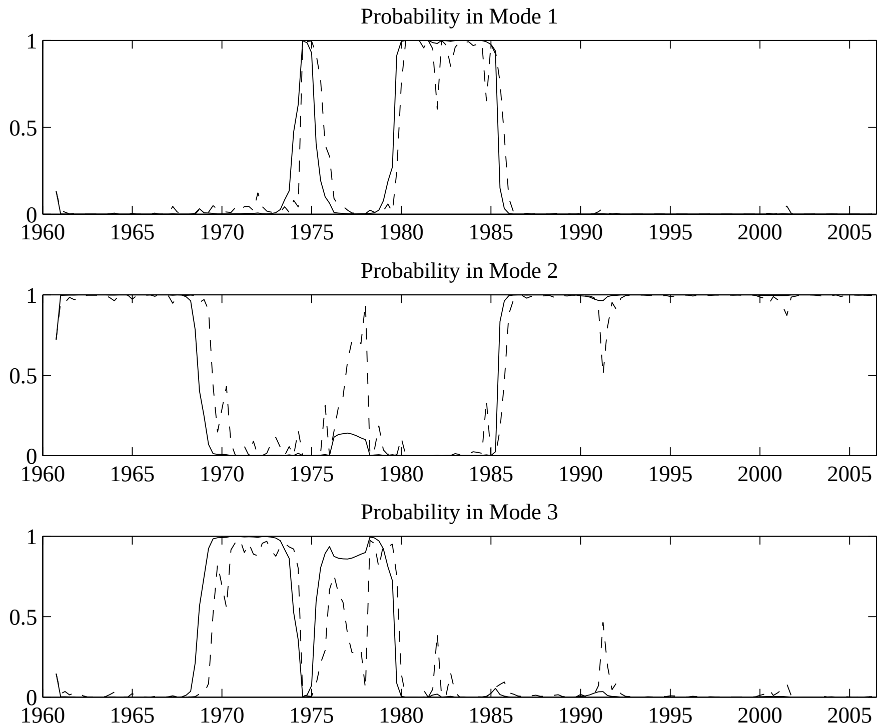

Figure 4.4: Estimated probabilities of being the different modes. Smoothed (full-sample) inference is shown with solid lines, while filtered (one-sided) inference is shown with dashed lines.

in figure 4.6. Again we consider 10,000 simulations of 50 periods, and plot the mean and median responses along with 30% probability bands and the corresponding optimal responses for the constant-coefficient model. [14] The distribution of the impulse response for observable and unobservable modes are again similar (note that the vertical scale varies from panel to panel), with generally more diffuse distributions in the unobservable case.

Model uncertainty leads to a change in the nature of policy also for the forward-looking model. Compared to the constant-coefficient case, most of the mass of the distribution of impulse responses is further away from zero, consistent with larger effects of the shocks. The effects are also more persistent. For example, both shocks have a much more persistent effect on inflation. In terms of the optimal policy responses, the instrument rate responses are more aggressive under our parameter uncertainty, especially for shocks to inflation. This is counter to the standard Brainard results of less aggressive response under uncertainty, although again our exercise here is fairly different. At least part of the difference in policy may be explained by the fact that inflation has more intrinsic persistence than the constant coefficient case (ω_f is lower) in two of the three modes, making policy

[14] Again, the shocks are $\varepsilon_{\pi 0} = 1$ and $\varepsilon_{y0} = 1$, respectively, so the shocks to the inflation and output-gap *equations* in period 0 are mode-dependent and equal to $c_{\pi j}$ and c_{yj} ($j = 1, 2, 3$), respectively. The distribution of modes in period 0 (and thereby all periods) is again the stationary distribution.

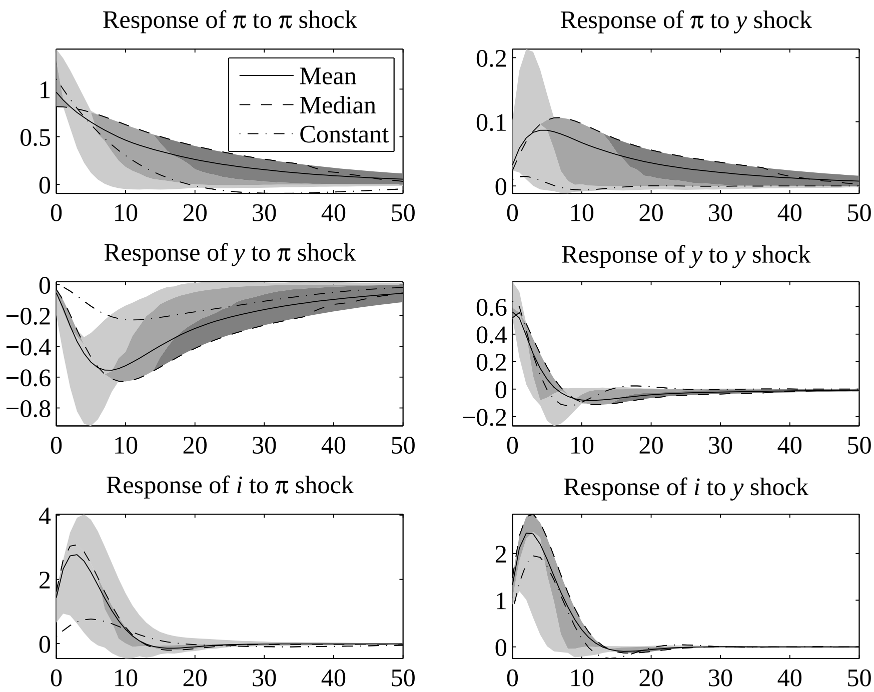

Figure 4.5: Unconditional impulse responses to shocks under the optimal policy for the mode-dependent Lindé model. Solid lines: Mean responses. Dark/medium/light grey bands: 30/60/90% probability bands. Dashed lines: Optimal responses for the constant-coefficient model.

more reactive.

Once again, the distribution of the impulse responses is asymmetric, with the mean responses quite different from the median responses. As in the backward-looking model above, this is perhaps most noticeable for the inflation responses, where the center of the distribution lies below the constant-coefficient case but there is a relatively large right tail showing more significant and persistent responses. Interestingly, the policy responses do die out much more quickly in the forward-looking model. The distribution of policy responses is tightly centered on zero after roughly 20 quarters, while in the backward-looking case even after 50 quarters there is still some spread in the distribution. Thus it appears that expectations may play an important role in stabilizing the economy under model uncertainty. We next turn to an example which highlights the role of expectations even more.

4.3 Uncertainty about whether the model is forward- or backward-looking

While the previous examples can be interpreted as either model or parameter uncertainty within a given class of structural models, we now turn to an example which more clearly captures uncertainty between classes of structural models. The degree to which inflation, in particular, is forward-looking

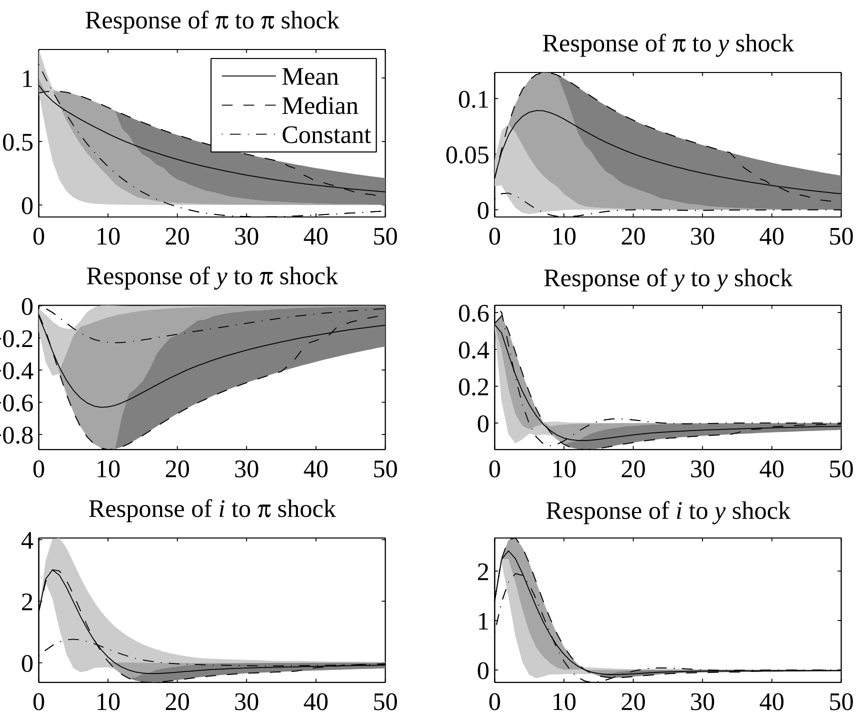

Figure 4.6: Unconditional impulse responses to shocks under the optimal policy for the unobservable-mode version of the Lindé model. Solid lines: Mean responses. Dark/medium/light grey bands: 30/60/90% probability bands. Dashed lines: Optimal responses for constant coefficients.

is a matter of much interest for policy and is subject to some debate. We also saw above that uncertainty about forward-looking components is a key element of the uncertainty about the Lindé model. In this section we focus more attention on this aspect, by considering a restricted version of that model. In particular, we consider a two-mode MJLQ model where one mode has forward- and backward-looking elements, while the other is backward-looking only. Thus we specify that mode 1 is unrestricted, while in mode 2 we restrict $\omega_f = \beta_f = 0$, so that the mode is backward-looking. Thus the model in mode 2 is similar to the RS model, albeit with fewer lags. We then re-estimate the model as in the previous case, using the same data set and the same priors for all unrestricted parameters.

Table 4.5 reports our estimates, with the estimates from the constant-coefficient version of the model (as above) for comparison. Here we see that apart from the forward-looking terms (which of course are restricted) the variation in the other parameters across the modes is relatively minor. There are some differences in the estimated policy rules, but relatively little change across modes in the other structural coefficients. The estimated transition matrix P and its implied stationary

Parameter	Constant	Mode 1	Mode 2
ω_f	0.5164	0.3272	0
γ	0.0034	0.0580	0.0432
β_f	0.4484	0.4801	0
β_r	0.0073	0.0114	0.0380
β_y	1.1902	1.5308	1.2538
ρ_1	0.8216	1.2079	1.2314
ρ_2	0.0560	0.9228	0.7127
γ_π	1.5538	0.7430	1.2845
γ_y	0.9777	0.1094	−0.3577
c_π	0.5920	1.0621	0.8301
c_y	0.3753	0.5080	0.5769
c_i	1.0412	1.7187	0.3848

Table 4.5: Estimates of the constant-coefficient and a restricted two-mode Lindé model.

Mode	π_{t-1}	y_{t-1}	y_{t-2}	i_{t-1}	$\varepsilon_{\pi t}$	ε_{yt}	$\Xi_{\pi,t-1}$	$\Xi_{y,t-1}$
Constant	0.1738	0.9394	−0.2112	0.7623	0.2128	0.7559	0.0011	0.0252
Mode 1	1.0237	0.8752	−0.2191	0.7159	1.6163	1.0298	0.0034	0.0303
Mode 2	2.0878	2.7114	−0.6116	0.5955	1.7331	1.5247	0.0000	0.0001
Unobservable	1.9893	2.3434	−0.5343	0.6228	1.8991	1.4685	0.0001	0.0031

Table 4.6: Optimal policy functions of the constant-coefficient and constrained two-mode Lindé model for observable and unobservable modes.

distribution $\bar{p}$ are given by

$$P = \begin{bmatrix} 0.9579 & 0.0421 \\ 0.0169 & 0.9831 \end{bmatrix}, \qquad \bar{p} = \begin{bmatrix} 0.2869 \\ 0.7131 \end{bmatrix}.$$

Thus mode 2 is the most persistent and has the largest mass in the invariant distribution. This is consistent with our estimation of the modes as shown in figure 4.7. Again, the plots show both the smoothed and filtered estimates. On the whole, our results here are similar to the unconstrained 3 mode model above. Mode 2, the backward-looking model mode, was experienced the most throughout much of the sample, holding for 1961-1968 and then with near certainty continually since 1985. The forward-looking model held in periods of rapid changes in inflation, holding for both the run-ups in inflation in the early and late 1970s and the disinflationary period of the early 1980s. During periods of relative tranquility, such as the Greenspan era, the backward-looking model fits the data the best.

We again solve for the optimal policy function, using the same methods and specification as above. The optimal policy functions for the constant-coefficient, observable-modes, and unobservable-modes cases are given in table 4.6. In figure 4.8, we plot the distribution of the impulse responses of inflation, the output gap, and the instrument rate to the two structural shocks in the model for the unobservable-modes case. Again we consider 10,000 simulations of 50 periods, and plot the mean responses and the corresponding optimal responses for the constant-coefficient model. However we do not plot the distribution of responses, as it gives much the same picture as in the unconstrained

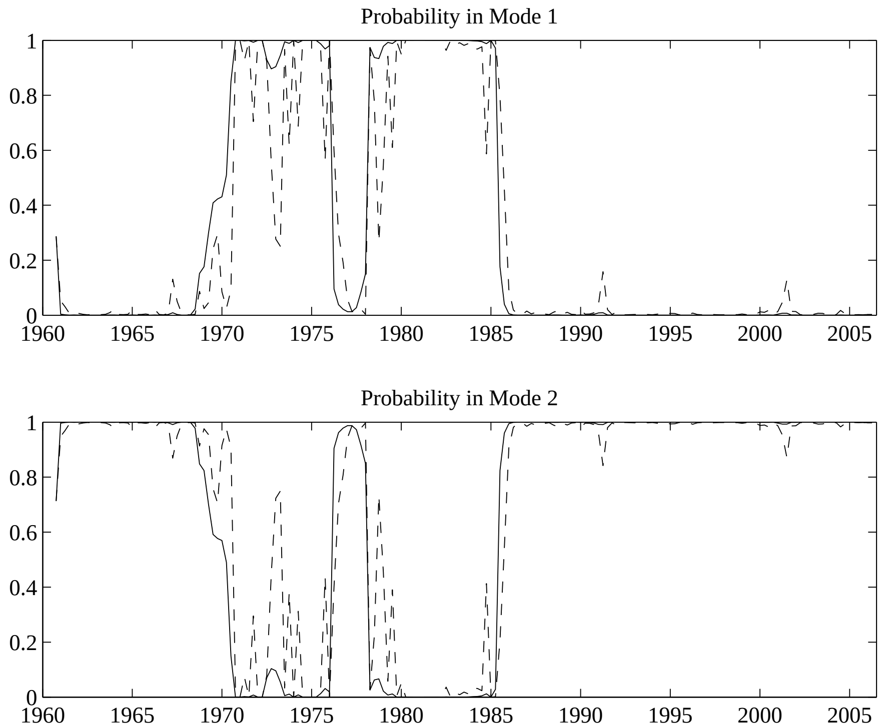

Figure 4.7: Estimated probabilities of being the different modes. Smoothed (full-sample) inference is shown with solid lines, while filtered (one-sided) inference is shown with dashed lines.

Lindé model. Instead, we now plot two different realizations of the responses, one in which mode 1 is realized throughout the 50 sample periods and one in which mode 2 is realized throughout.

The effects of model uncertainty on policy are very much as in the unconstrained model above. Compared to the constant-coefficient case, the mean impulse responses are consistent with larger effects of the shocks which are also longer lasting. In terms of the optimal policy responses, the instrument rate responses are again more aggressive with model uncertainty, especially with regard to the inflation shocks. What is also interesting to compare in this case is the differences in responses in the two modes. Interestingly, for both shocks the policy responses are very similar in the forward- and backward-looking modes. However in the forward-looking mode 1, the effects on output and inflation are somewhat sharper and more immediate, with the effects of the shock dying out more quickly. In contrast, for the backward-looking mode 2 the shocks have much more persistent effects, particularly for inflation. This highlights the important role that expectations have in stabilizing the economy in response to shocks.

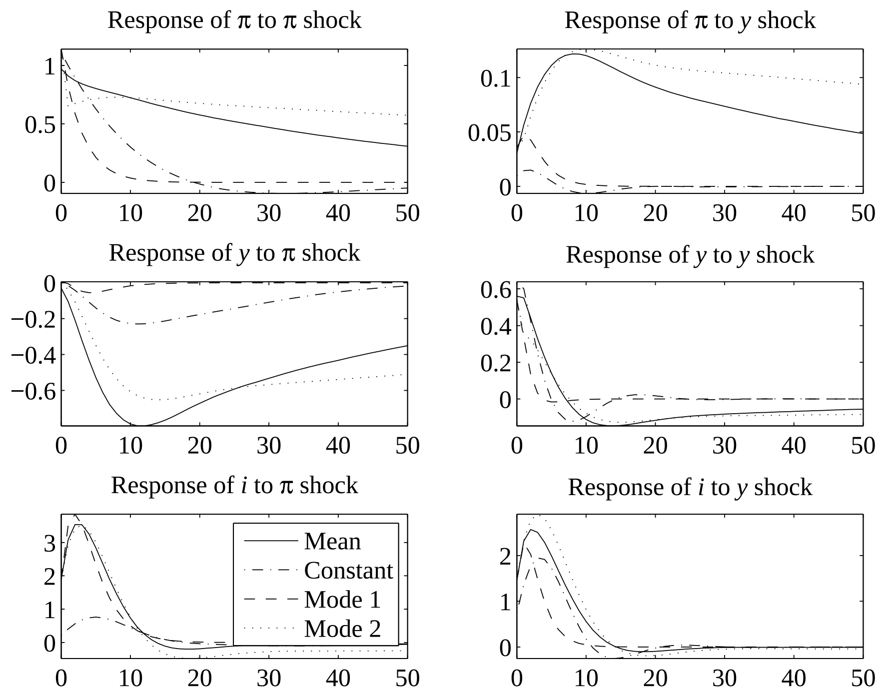

Figure 4.8: Unconditional impulse responses to shocks under the optimal policy for the unobservable-modes version of the constrained two-mode Lindé model. Solid lines: Mean responses. Dot-dashed lines: Optimal responses for the constant-coefficient model. Dashed lines: optimal responses when mode 1 is realized throughout. Dotted lines: optimal responses when mode 2 is realized throughout.

5 Arbitrary instrument rules and optimal restricted instrument rules

There is a large literature on restricted or "simple" instrument rules in policy. Here we show how to implement and optimize over simple rules in our setting. In particular, we derive the dynamics of the system, including the distribution of forecasts of relevant future variables, for arbitrary time-invariant instrument rules. We also solve for optimal simple rules in one of our estimated examples from above. For simplicity, we focus here on the observable modes case only.

5.1 Instrument rules

Consider an arbitrary time-invariant instrument rule of the form

$$i_t = F_{Xj_t} X_t + F_{xj_t} x_t \qquad (j_t \in N_j), \tag{5.1}$$

to be combined with (2.1) and (2.2). If $F_{xj_t} \equiv 0$, (5.1) is an *explicit* instrument rule; that is, the instrument responds to predetermined variables only.[15] If $F_{xj_t} \not\equiv 0$ ($F_{xj_t} \neq 0$ for some mode j_t with positive probability), it is an *implicit* instrument rule; that is, the instrument depends also on forward-looking variables. In the latter case, there is a simultaneity problem, in that the instrument and the forward-looking variables are simultaneously determined. Thus, an implicit instrument rule can be interpreted as an equilibrium condition. As discussed in Svensson [39] and Svensson and Woodford [43], the implementation of an implicit instrument rule is problematic, since in practice a central bank can literally only respond to predetermined variables.[16] We disregard these problems here, and consider (5.1) as just another equilibrium condition added to equations (2.1) and (2.2).

Although we explicitly only deal with instrument rules here, our method generalizes to arbitrary policy rules, including targeting rules (Svensson and Woodford [43]). Indeed, we could consider any arbitrary policy rule, including targeting rules, of the form

$$\mathrm{E}_t H_{3j_{t+1}} x_{t+1} = A_{31j_t} \tilde{X}_t + A_{32j_t} x_t + B_{3j_t} i_t,$$

where H_{3j}, A_{31j}, A_{32j}, and B_{3j} are potentially mode-dependent matrices of the appropriate dimension (in particular, having n_i rows and giving rise to n_i independent equations, which is required to determine the instruments in each period).

We assume that combining (5.1) with (2.1) and (2.2) results in a unique solution.[17] This solution can be written

$$\begin{aligned}
x_t &= G_{j_t} X_t, \\
i_t &= (F_{Xj_t} + F_{xj_t} G_{j_t}) X_t \;\equiv\; F_{j_t} X_t. \\
X_{t+1} &= (A_{11j_{t+1}} + A_{12j_{t+1}} G_{j_t} + B_{1j_{t+1}} F_{j_t}) X_t + C_{j_{t+1}} \varepsilon_{t+1} \;\equiv\; M_{j_t j_{t+1}} X_t + C_{j_{t+1}} \varepsilon_{t+1},
\end{aligned}$$

where the matrices G_{j_t} ($j_t \in N_j$) are endogenously determined. This solutions will give rise to a probability distribution of $X_{t+\tau}$, $x_{t+\tau}$, and $i_{t+\tau}$ ($\tau \geq 0$) conditional on X_t and j_t. This solution will be associated with a value function for the original period loss function,

$$X_t' V_{j_t} X_t + w_{j_t}.$$

In appendix G we develop an iterative algorithm to solve for the equilibrium and the value of the loss.

For a given restricted class $\mathcal{F}$ of instrument rules, we can consider the optimal restricted (time-invariant) instrument rule $\hat{F}$, which minimizes an intertemporal loss function. This intertemporal

[15] Note that policy functions are explicit instrument rules.

[16] In practice, because of a complex and systematic decision process (Brash [7], Sims [34], Svensson [37]), the information modern central banks respond to is at least a few days old, and most of the information is one or several months old.

[17] Farmer, Waggoner, and Zha [16] provide conditions for uniqueness in models like ours, and develop an algorithm that solves for indeterminate equilibria as well. They find that our results agree with theirs when the equilibrium is unique.

loss function could be the conditional loss in a given period, say period 0,

$$\hat{F} \equiv \arg\min_{F\in\mathcal{F}} \{\tilde{X}_0' V_{j_0}(F)\tilde{X}_0 + w_{j_0}(F)\},$$

where the notation takes into account that $V_{j_0}(F)$ and $w_{j_0}(F)$ depend on $F \in \mathcal{F}$. This would make the optimal restricted time-invariant instrument rule depend on $\tilde{X}_0$, j_0, and the covariance matrices $\tilde{C}_j\tilde{C}_j'$ of the shocks $\tilde{C}_j\varepsilon_{t+1}$ to $\tilde{X}_{t+1}$ in mode $j \in N_j$. Alternatively, the intertemporal loss function could be the unconditional mean of the period loss function:

$$\hat{F} = \arg\min_{F\in\mathcal{F}} \mathrm{E}[L_t].$$

Note that

$$\mathrm{E}[L_t] = (1-\delta)\mathrm{E}[\tilde{X}_t' V_{j_t}(F)\tilde{X}_t + w_{j_t}(F)].$$

Furthermore, the unconditional and conditional intertemporal loss are approximately the same when the intertemporal loss is scaled by $1-\delta$ and δ is close to one,

$$\lim_{\delta\to 1} \mathrm{E}_t \sum_{\tau=0}^{\infty}(1-\delta)\delta^{\tau} L_{t+\tau} = \mathrm{E}[L_t] = \lim_{\delta\to 1}(1-\delta)\mathrm{E}[w_{j_t}] = \mathrm{E}[\mathrm{tr}(V_{j_t}\tilde{C}_{j_t}\tilde{C}_{j_t}')] = \sum_j \bar{p}_j \mathrm{tr}(V_j\tilde{C}_j\tilde{C}_j'),$$

where we recall that $\bar{p} = (\bar{p}_1, ..., \bar{p}_{n_j})'$ is the stationary distribution of modes.

5.2 Optimal Taylor-type instrument rules in a forward-looking model

We now apply the methods outlined above to derive optimal Taylor-type instrument rules in the estimated forward-looking model from section 4.2 above. In particular, we consider simple implicit instrument rules of the general form (disregarding the implementation problems mentioned above):

$$i_t = f_{ij_t} i_{t-1} + f_{\pi j_t}\pi_t + f_{yj_t} y_t. \tag{5.2}$$

This is a Taylor rule with interest-rate smoothing, whose coefficients may depend on the mode j_t in period t. As special cases, we consider mode-independent Taylor rules, where the coefficients are constrained to be the same in all modes, and original Taylor rules without the smoothing coefficient f_i. We use the unconditional mean of the period loss, $\mathrm{E}[L_t]$, as the intertemporal loss function.

The results are summarized in table 5.1, where we report the optimal response coefficients of the different forms of the instrument rules for the constant-coefficient and MJLQ versions of the model. We find that the optimal Taylor-type rules that are constrained to have the same responses in all modes are more aggressive in the MJLQ model than in the constant-coefficient model. This is in line with the impulse responses for the optimal policy shown in figure 4.5 above, where we found that the optimal policy in the MJLQ model had on average a substantially more aggressive inflation response and a slightly more aggressive output-gap response than in the constant-coefficient model. Similar conclusions apply for both the original and smoothed Taylor rules. This increased aggressiveness

Mode	i_{t-1}	π_t	y_t	Loss
Constant-coefficient model				
Optimal policy function				8.27
-	-	1.88	1.68	11.06
-	0.90	0.44	0.75	9.34
MJLQ model				
Optimal policy function				17.39
All modes	-	3.87	3.35	20.89
Mode 1	-	3.20	-0.30	
Mode 2	-	4.16	3.83	20.01
Mode 3	-	1.96	1.10	
All modes	0.73	1.90	1.93	17.96
Mode 1	0.86	1.49	1.04	
Mode 2	0.65	2.24	2.29	17.52
Mode 3	0.76	0.78	0.99	

Table 5.1: Optimal Taylor-type instrument rules for the estimated three-mode Lindé model.

is further illustrated in figure 5.1. The figure shows the loss in the constant-coefficient and MJLQ models for mode-independent original and smoothed Taylor rules (with the smoothing coefficient fixed at $f_i = 0.8$). For both smoothed and original Taylor rules, the loss function is more sensitive to variations in the inflation response coefficient of the policy rule than the output gap response. For both kinds of rules, performance in the MJLQ model is enhanced by more aggressive responses.

These results suggest that constraining the rules to react in the same way in all modes may push the optimal simple rules towards more aggressive responses. To see whether this overall aggressiveness is affected by averaging across modes, we also consider mode-dependent original and smoothed Taylor rules, which are reported in table 5.1. There we see that there is significant variation in the responses across modes. For example, the most aggressive policy responses are in mode 2, which recall is the most backward-looking mode. However, as the table shows, tailoring the coefficients of either an original or smoothed rule to the different modes reduces losses by relatively little. In contrast, smoothing the policy response has significant effects, driving the losses in the MJLQ model down to nearly the optimal level. Above we saw that the effects of uncertainty, captured by comparison of the constant coefficient model to the MJLQ model, led to more aggressive optimal policy. Those results are reinforced here using a different metric: the coefficients of an instrument rule rather than the magnitude of the policy responses to shocks.

6 Conclusions

This paper demonstrates that the Markov jump-linear-quadratic (MJLQ) framework is a very flexible and powerful tool for the analysis and determination of optimal policy under model uncertainty. It provides a very tractable way of handling the absence of certainty equivalence that is an important aspect of model uncertainty. Our approach builds on the control-theory literature, for instance,

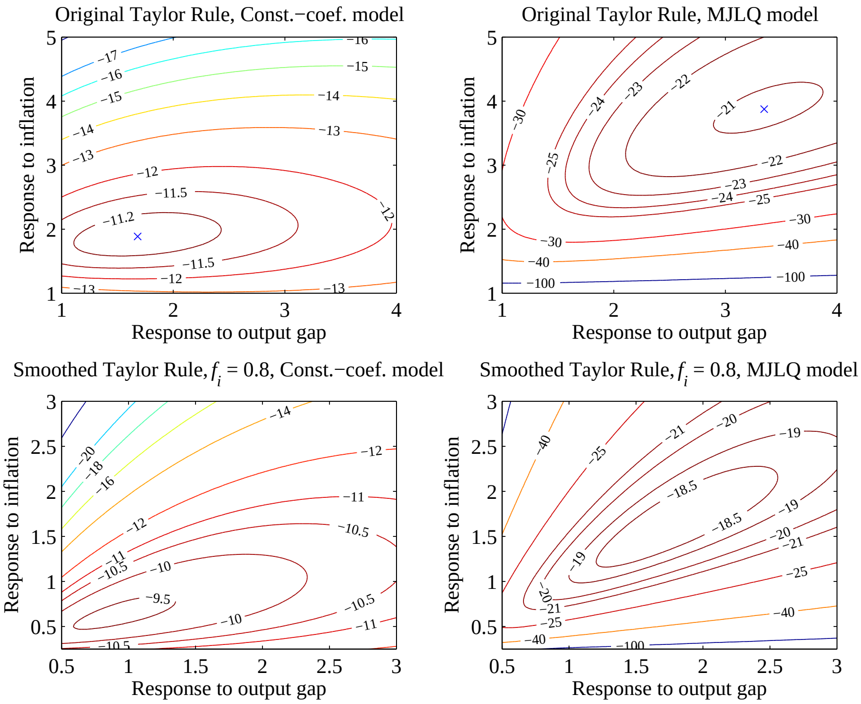

Figure 5.1: Contours of the loss function for the Lindé model under mode-independent Taylor-type instrument rules. Left column: Constant-coefficient model. Right column: MJLQ model. Top row: Original Taylor rules. Bottom row: Smoothed Taylor rules with $f_i = 0.8$.

Costa, Fragoso, and Marques [11], which has explored many properties of the MJLQ framework. That literature uses recursive methods and does not consider forward-looking variables. However the forward-looking variables characteristic of rational expectations make the models nonrecursive. We show that the recursive saddlepoint method of Marcet and Marimon [29] can be applied to this problem which allows us to use recursive methods, and hence to solve relatively general models.

We show that our framework can incorporate a large variety of different configurations of uncertainty. We provide algorithms to derive the optimal policy and value functions. We apply the framework to regime-switching variants of two empirical models of the US economy, the backward-looking model of Rudebusch and Svensson [33] and the forward-looking New Keynesian model of Lindé [27]. We also show how the dynamics of the model can be specified for arbitrary instrument rules, and how to optimize over restricted instrument rules. Finally, we show how the framework can be adapted to a situation with unobservable modes, arguably the most realistic situation for policy. In the examples we study, we find some substantial deviations from certainty equivalence. In some cases, we find support for the common intuition that uncertainty should make policy more cautious. But this is not a general result, and depends on the nature of the uncertainty

The MJLQ framework makes it possible to provide advice on optimal monetary policy for a large variety of different configurations of model uncertainty. While the particular examples we study in this paper are informative, they are only a small sample of the applications which can be analyzed with our approach. The framework also makes it possible to incorporate different kinds of central-bank judgment—information, knowledge, and views outside the scope of a particular model—about the kind and degree of model uncertainty. Furthermore, the framework can incorporate the kind of central-bank judgment about additive future deviations—add factors—that is discussed in Svensson [39] and Svensson and Tetlow [41]. Some additional natural applications would embed the different specifications of fully specified dynamic stochastic general equilibrium models as modes in the MJLQ setting. We could thus incorporate uncertainty about the structure of the economy, such as different forms of price or wage setting (as discussed in Levin, Onatski, Williams, and Williams [26]). Alternative specifications could also capture uncertainty about the low-frequency behavior of the key driving processes, which could describe potential productivity slowdowns (as in, for example, Kahn and Rich [23]) or moderations in overall volatility (as in McConnell and Perez-Quiros [30] and Stock and Watson [36]). Our approach clearly can capture a wide variety of different types of uncertainty which are relevant for policy.

Overall, our results point to some important changes from approaches considering additive uncertainty. In the "mean forecast targeting" applications in Svensson [39] and Svensson and Tetlow [41], certainty equivalence is preserved, since the uncertainty is restricted to additive future stochastic deviations in the model's equations. With certainty equivalence, only the means of future variables matter for policy, and optimal policy can be derived as if there were no uncertainty about those means. Furthermore, the optimal mean projection of future target variables and the instrument can be calculated in one step, and those projections—including the optimal mean instrument path— are the natural objects for policy discussion. There is no need to use recursive methods, and there is no need to specify the optimal policy function for the policy makers (the explicit policy function is also a high-dimensional vector that is not easy to interpret). Instead, the policy discussion can be conducted with the help of computer-generated graphs of projections of the target variables and the instrument under alternative assumptions, weights in the monetary-policy loss function, and central-bank judgments.

In the absence of certainty equivalence, mean forecast targeting is in principle no longer sufficient. The whole distribution of future target variables matters for policy, and the optimal instrument decision should in principle take this into account. The optimal policy plan should be chosen such that the whole distribution, rather than the mean, of the future target variables "looks good." The central bank should engage in "distribution forecast targeting" rather than mean forecast targeting. The application of the MJLQ framework in this paper to model uncertainty and certainty non-equivalence indicates that recursive methods and the explicit policy function are relatively more useful for the derivation of the optimal policy than under certainty equivalence, perhaps even necessary. Still, the resulting distributions of future target variables and instruments under alternative assumptions can conveniently be illustrated and presented to policy makers in the form of graphs, albeit graphs of distributions rather than of means.

Appendix

A Incorporating central-bank judgment

In order to incorporate (additive) central-bank judgment as in Svensson [39], consider the model

$$X_{t+1} = A_{11j_{t+1}} X_t + A_{12j_{t+1}} x_t + B_{1j_{t+1}} i_t + C_{t+1} \tilde{z}_{t+1}, \tag{A.1}$$

$$\mathrm{E}_t H_{j_{t+1}} x_{t+1} = A_{21j_t} X_t + A_{22j_t} x_t + B_{2j_t} i_t, \tag{A.2}$$

where $\tilde{z}_t$, the (additive) *deviation*, is a an exogenous $n_{\tilde{z}}$-vector stochastic process. Assume that $\tilde{z}_t$ satisfies

$$\tilde{z}_{t+1} = \varepsilon_{t+1} + \sum_{s=1}^{T} \varepsilon_{t+1,t+1-s}$$

for given $T \geq 0$, where $(\varepsilon_t', \varepsilon^{t\prime})' \equiv (\varepsilon_t', \varepsilon_{t+1,t}', ..., \varepsilon_{t+T,t}')'$ is a zero-mean i.i.d. random $(T+1)n_{\tilde{z}}$-vector realized in the beginning of period t and called the innovation in period t. For $T = 0$, $\tilde{z}_{t+1} = \varepsilon_{t+1}$ is a simple i.i.d. disturbance. For $T > 0$, the deviation is a version of a moving-average process.

The dynamics of the deviation can be written

$$\begin{bmatrix} \tilde{z}_{t+1} \\ \tilde{z}^{t+1} \end{bmatrix} = A_{\tilde{z}} \begin{bmatrix} \tilde{z}_t \\ \tilde{z}^t \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon^{t+1} \end{bmatrix},$$

where $\tilde{z}^t \equiv (\mathrm{E}_t \tilde{z}_{t+1}', \mathrm{E}_t \tilde{z}_{t+2}', ..., \mathrm{E}_t \tilde{z}_{t+T})'$ can be interpreted as the central bank's (additive) *judgment* in period t and the $(T+1)n_{\tilde{z}} \times (T+1)n_{\tilde{z}}$ matrix $A_{\tilde{z}}$ is defined as

$$A_{\tilde{z}} \equiv \begin{bmatrix} 0_{n_{\tilde{z}} \times n_{\tilde{z}}} & I_{n_{\tilde{z}}} & 0_{n_{\tilde{z}} \times (T-1)n_{\tilde{z}}} \\ 0_{(T-1)n_{\tilde{z}} \times n_{\tilde{z}}} & 0_{(T-1)n_{\tilde{z}} \times n_{\tilde{z}}} & I_{(T-1)n_{\tilde{z}}} \\ 0_{n_{\tilde{z}} \times n_{\tilde{z}}} & 0_{n_{\tilde{z}} \times n_{\tilde{z}}} & 0_{n_{\tilde{z}} \times (T-1)n_{\tilde{z}}} \end{bmatrix} \equiv \begin{bmatrix} 0 & A_{\tilde{z}21} \\ 0 & A_{\tilde{z}22} \end{bmatrix};$$

in the second identity $A_{\tilde{z}}$ is partitioned conformably with $\tilde{z}_t$ and $\tilde{z}^t$. Hence $\tilde{z}^t$ is the central bank's mean projection of future deviations, and ε^t can be interpreted as the new information the central bank receives in period t about those future deviations.[18]

It follows that the model can be written in the state-space form (2.1) and (2.2) as

$$\begin{bmatrix} X_{t+1} \\ \tilde{z}_{t+1} \\ \tilde{z}^{t+1} \end{bmatrix} = \hat{A}_{11j_{t+1}} \begin{bmatrix} X_t \\ \tilde{z}_t \\ \tilde{z}^t \end{bmatrix} + \hat{A}_{12j_{t+1}} x_t + \hat{B}_{1j_{t+1}} i_t + \hat{C}_{j_{t+1}} \begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon^{t+1} \end{bmatrix},$$

$$\mathrm{E}_t H_{j_{t+1}} x_{t+1} = \hat{A}_{21j_t} \begin{bmatrix} X_t \\ \tilde{z}_t \\ \tilde{z}^t \end{bmatrix} + A_{22j_t} x_t + B_{2j_t} i_t,$$

[18] The graphs in Svensson [39] can be seen as impulse responses to ε^t.

where

$$\hat{A}_{11j_{t+1}} \equiv \begin{bmatrix} A_{11j_{t+1}} & 0 & C_{j_{t+1}}A_{\tilde{z}21} \\ 0 & 0 & A_{\tilde{z}21} \\ 0 & 0 & A_{\tilde{z}22} \end{bmatrix}, \quad \hat{A}_{21j_{t+1}} \equiv \begin{bmatrix} A_{21j_{t+1}} \\ 0 \\ 0 \end{bmatrix},$$

$$\hat{B}_{1j_{t+1}} \equiv \begin{bmatrix} B_{1j_{t+1}} \\ 0 \\ 0 \end{bmatrix}, \quad \hat{C}_{j_{t+1}} \equiv \begin{bmatrix} C_{j_{t+1}} & 0 \\ I_{n_{\tilde{z}}} & 0 \\ 0 & I_{n_{\tilde{z}}} \end{bmatrix},$$

and the new predetermined variables are $(X_t', \tilde{z}_t', \tilde{z}^{t\,\prime})'$.

B An algorithm for the case of observable modes

Consider the dual saddlepoint problem of (2.13) in period t, subject to (2.11), (2.15), (2.16) and $\tilde{X}_t$ given. Let us use the notation $Z_t = Z_{j_t}$ for any matrix Z that is a function of the mode j_t, and let the matrix $\tilde{W}_t = \tilde{W}_{j_t}$ in (2.12) be partitioned conformably with $\tilde{X}_t$ and $\tilde{\imath}_t$ as

$$\tilde{W}_t \equiv \begin{bmatrix} Q_t & N_t \\ N_t' & R_t \end{bmatrix}.$$

We use that the value function for the dual problem will be quadratic and can be written

$$\tilde{X}_t'\tilde{V}_t\tilde{X}_t + \tilde{w}_t,$$

where $\tilde{V}_t$ is a matrix and $\tilde{w}_t$ a scalar. It will satisfy the Bellman equation

$$\tilde{X}_t'\tilde{V}_t\tilde{X}_t + \tilde{w}_t = \max_{\gamma_t} \min_{(x_t, i_t)} \left\{ \tilde{X}_t'Q_t\tilde{X}_t + 2\tilde{X}_t'N_t\tilde{\imath}_t + \tilde{\imath}_t'R_t\tilde{\imath}_t + \delta \mathrm{E}_t(\tilde{X}_{t+1}'\tilde{V}_{t+1}\tilde{X}_{t+1} + \tilde{w}_{t+1}) \right\},$$

where $\tilde{X}_{t+1} \equiv (X_t', \Xi_{t-1}')'$ and E_t refers to the expectations conditional on $\tilde{X}_t$ and j_t.

The first-order condition with respect to $\tilde{\imath}_t$ is

$$\tilde{X}_t'N_t + \tilde{\imath}_t'R_t + \delta\tilde{X}_t'\mathrm{E}_t\tilde{A}_{t+1}'\tilde{V}_{t+1}\tilde{B}_{t+1} + \delta\tilde{\imath}_t'\mathrm{E}_t\tilde{B}_{t+1}'\tilde{V}_{t+1}\tilde{B}_{t+1} = 0,$$

which can be written

$$J_t\tilde{\imath}_t + K_t\tilde{X}_t = 0,$$

where

$$J_t \equiv R_t + \delta\mathrm{E}_t\tilde{B}_{t+1}'\tilde{V}_{t+1}\tilde{B}_{t+1}, \tag{B.1}$$

$$K_t \equiv N_t' + \delta\mathrm{E}_t\tilde{B}_{t+1}'\tilde{V}_{t+1}\tilde{A}_{t+1}. \tag{B.2}$$

This leads to the optimal policy function

$$\tilde{\imath}_t = F_t\tilde{X}_t, \tag{B.3}$$

where

$$F_t \equiv -J_t^{-1}K_t. \tag{B.4}$$

Furthermore, the value function satisfies

$$\tilde{X}_t'\tilde{V}_t\tilde{X}_t \equiv \tilde{X}_t'Q_t\tilde{X}_t + 2\tilde{X}_t'N_tF_t\tilde{X}_t + \tilde{X}_t'F_t'R_tF_t\tilde{X}_t + \delta\tilde{X}_t'\mathrm{E}_t[(\tilde{A}_{t+1}' + F_t'\tilde{B}_{t+1}')\tilde{V}_{t+1}(\tilde{A}_{t+1} + \tilde{B}_{t+1}F_t)]\tilde{X}_t.$$

This implies

$$\tilde{V}_t = Q_t + N_tF_t + F_t'N_t' + F_t'R_tF_t + \delta\mathrm{E}_t[(\tilde{A}_{t+1}' + F_t'\tilde{B}_{t+1}')\tilde{V}_{t+1}(\tilde{A}_{t+1} + \tilde{B}_{t+1}F_t)],$$

which can be simplified to the Riccati equation

$$\tilde{V}_t = Q_t + \delta\mathrm{E}_t\tilde{A}_{t+1}'\tilde{V}_{t+1}\tilde{A}_{t+1} - K_t'J_t^{-1}K_t. \tag{B.5}$$

Equations (B.1), (B.2), and (B.5) show how $\tilde{V}_{t+1} = \tilde{V}_{j_{t+1}}$ for $j_{t+1} = 1, ..., n_j$ is mapped into $\tilde{V}_t = \tilde{V}_{j_t}$ for $j_t = 1, ..., n_j$.

Iteration backwards of (B.4) and (B.5) from any constant positive semidefinite matrix $\tilde{V}$ should converge to stationary matrices functions F_j and $\tilde{V}_j$ $(j \in N_j)$, where $\tilde{V}_j$ satisfies the Riccati equation (B.5) with (B.1) and (B.2).

Taking account of the finite number of modes, we have

$$\begin{aligned} F_j &\equiv -J_j^{-1}K_j \\ J_j &\equiv R_j + \delta\sum_{k=1}^{n_j} P_{jk}\tilde{B}_k'\tilde{V}_k\tilde{B}_k \\ K_j &\equiv N_j' + \delta\sum_{k=1}^{n_j} P_{jk}\tilde{B}_k'\tilde{V}_k\tilde{A}_k, \\ \tilde{V}_j &= Q_j + \delta\sum_{k=1}^{n_j} P_{jk}\tilde{A}_k'\tilde{V}_k\tilde{A}_k - K_j'J_j^{-1}K_j \qquad (j \in N_j), \end{aligned} \tag{B.6}$$

where P_{jk} is the transition probability from $j_t = j$ to $j_{t+1} = k$.

The scalars $\tilde{w}_j$ solve the equations

$$\tilde{w}_j = \delta\sum_k P_{jk}[\mathrm{tr}(\tilde{V}_k\tilde{C}_k\tilde{C}_k') + \tilde{w}_k].$$

Thus determining the optimal policy function (B.3) reduces to solving a system of coupled algebraic Riccati equations (B.6). In order to solve this system numerically, we adapt the algorithm of do Val, Geromel, and Costa [15]. In a very similar problem, they show how the coupled Riccati equations can be uncoupled for numerical solution.[19]

The algorithm consists of the following steps:

1. Define $\hat{A}_j = \sqrt{P_{jj}}\tilde{A}_j,\ \hat{B}_j = \sqrt{P_{jj}}\tilde{B}_j$ and initialize $\tilde{V}_j^0 = 0,\ j = 1, \ldots, n_j$.

[19] In their problem, the matrices A and B next period are known in the current period, so the averaging in the Riccati equation is only over the V_j matrices.

2. Then at each iteration $l = 0, 1, \dots,$ for each j define:

$$\begin{aligned}
\hat{Q}_j &= Q_j + \delta \sum_{k \neq j} P_{jk} \tilde{A}_k' \tilde{V}_k^l \tilde{A}_k \\
\hat{R}_j &= R_j + \delta \sum_{k \neq j} P_{jk} \tilde{B}_k' \tilde{V}_k^l \tilde{B}_k \\
\hat{N}_j &= N_j + \delta \sum_{k \neq j} P_{jk} \tilde{A}_k' \tilde{V}_k^l \tilde{B}_k.
\end{aligned}$$

   Then for each j solve the standard Riccati equation for the problem with matrices $(\hat{A}_j, \hat{B}_j, \hat{Q}_j, \hat{R}_j, \hat{N}_j)$. Note that these are uncoupled since $\tilde{V}_k^l$ is known. Call the solution $\tilde{V}_j^{l+1}$.

3. Check $\sum_{j=1}^{n_j} \|\tilde{V}_j^{l+1} - \tilde{V}_j^l\|$. If this is lower then a tolerance, stop. Otherwise, return to step 2.

do Val, Geromel, and Costa [15] show that the sequence of matrices $\tilde{V}_j^l$ converges to the solution of (B.6) as $l \to \infty$. In order to understand the algorithm, recall that, in the standard linear-quadratic regulator (LQR) problem (Anderson, Hansen, McGrattan, and Sargent [1] and Ljungqvist and Sargent [28]), we have

$$\begin{aligned}
F &\equiv -J^{-1}K \\
J &\equiv R + \delta B'VB \\
K &\equiv N' + \delta B'VA, \\
V &= Q + \delta A'VA - K'J^{-1}K.
\end{aligned}$$

If we can redefine the matrices so the equations conform to the standard case, we can use the standard algorithm for the LQR problem to find F_j and V_j. The above definitions allow us to write

$$\begin{aligned}
F_j &\equiv -J_j^{-1}K_j, \\
J_j &\equiv \hat{R}_j + \delta \hat{B}_j' \tilde{V}_j \hat{B}_j, \\
K_j &\equiv \hat{N}_j' + \delta \hat{B}_j' \tilde{V}_j \hat{A}_j, \\
\tilde{V}_j &= \hat{Q}_j + \delta \hat{A}_j' \tilde{V}_j \hat{A}_j - K_j' J_j^{-1} K_j \qquad (j \in N_j),
\end{aligned}$$

so we can indeed use the standard algorithm.

C An algorithm for the case of unobservable modes

Our first task is to write the extended MJLQ system for the saddlepoint problem with unobservable modes. We suppose that we start with an initial period loss function (2.5) which has the form

$$L(X_t, x_t, i_t, j_t) \equiv \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix}' W_{j_t} \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix} \equiv \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix}' \begin{bmatrix} Q_{11j_t} & Q_{12j_t} & N_{1j_t} \\ Q_{12j_t}' & Q_{22j_t} & N_{2j_t} \\ N_{1j_t}' & N_{2j_t}' & R_{j_t} \end{bmatrix} \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix}.$$

The dual loss is

$$L(X_t, x_t, i_t, j_t) - \gamma_t' z_t + \Xi_{t-1}' \frac{1}{\delta} H_{j_t} x_t.$$

We now substitute in for x_t using

$$\begin{aligned} x_t = \tilde{x}(X_t, z_t, i_t, j_t) \;\equiv\; & - A_{22,j}^{-1} A_{21,j} X_t + A_{22,j}^{-1} z_t - A_{22,j}^{-1} B_{2,j} i_t \\ \equiv\; & A_{xX,j} X_t + A_{xz,j} z_t + A_{xi,j} i_t. \end{aligned} \tag{C.1}$$

After this substitution, we want to express the laws of motion and dual loss in terms of the expanded predetermined variables, $\tilde{X}_t = (X_t', \Xi_{t-1}')'$, and the expanded control variables, $\tilde{\imath}_t = (z_t', i_t', \gamma_t')'$. Suppressing time and mode subscripts on the right side for the time being (all are t and j_t, respectively, except $t-1$ on Ξ_{t-1}), we see that the dual loss can be written explicitly as

$$\tilde{L}(\tilde{X}_t, z_t, \gamma_t, i_t, j_t) \equiv$$

$$\begin{aligned} & X' \left(Q_{11} + A_{xX}' Q_{22} A_{xX} + 2A_{xX}' Q_{12}'\right) X + 2X' \left(N_1 + Q_{12} A_{xi} + A_{xX}' Q_{22} A_{xi} + A_{xX}' N_2\right) i \\ & + \; 2z'(A_{xz}' Q_{12}' + A_{xz}' Q_{22} A_{xX}) X + \Xi' \frac{1}{\delta} H A_{xX} X + \Xi' \frac{1}{\delta} H A_{xz} z + \Xi' \frac{1}{\delta} H A_{xi} i \\ & - \; \gamma' z + z'(A_{xz}' Q_{22} A_{xz}) z + i'(R + A_{xi}' Q_{22} A_{xi} + 2A_{xi}' N_2) i + 2z'(A_{xz}' N_2 + A_{xz}' Q_{22} A_{xi}) i. \end{aligned}$$

Thus we can write the dual loss

$$\tilde{L}(\tilde{X}_t, z_t, \gamma_t, i_t, j_t) \equiv \begin{bmatrix} \tilde{X}_t \\ \tilde{\imath}_t \end{bmatrix}' \begin{bmatrix} \tilde{Q}_j & \tilde{N}_j \\ \tilde{N}_j' & \tilde{R}_j \end{bmatrix} \begin{bmatrix} \tilde{X}_t \\ \tilde{\imath}_t \end{bmatrix},$$

where (again suppressing the j_t index)

$$\tilde{Q} \equiv \begin{bmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ \tilde{Q}_{12}' & 0 \end{bmatrix},$$

$$\begin{aligned} \tilde{Q}_{11} &\equiv Q_{11} + A_{xX}' Q_{22} A_{xX} + 2A_{xX}' Q_{12}', \\ \tilde{Q}_{12} &\equiv \frac{1}{2\delta} A_{xX}' H', \end{aligned}$$

$$\tilde{N} \equiv \begin{bmatrix} \tilde{N}_{11} & \tilde{N}_{12} & 0 \\ \tilde{N}_{21} & \tilde{N}_{22} & 0 \end{bmatrix},$$

$$\begin{aligned} \tilde{N}_{11} &\equiv Q_{12} A_{xz} + A_{xX}' Q_{22} A_{xz}, \\ \tilde{N}_{12} &\equiv N_1 + Q_{12} A_{xi} + A_{xX}' Q_{22} A_{xi} + A_{xX}' N_2, \\ \tilde{N}_{21} &\equiv \frac{1}{2\delta} H A_{xz}, \\ \tilde{N}_{22} &\equiv \frac{1}{2\delta} H A_{xi}, \end{aligned}$$

$$\tilde{R} \equiv \begin{bmatrix} \tilde{R}_{11} & \tilde{R}_{12} & \tilde{R}_{13} \\ \tilde{R}'_{12} & \tilde{R}_{22} & 0 \\ \tilde{R}'_{13} & 0 & 0 \end{bmatrix},$$

$$\begin{aligned}
\tilde{R}_{11} &\equiv A'_{xz} Q_{22} A_{xz}, \\
\tilde{R}_{12} &\equiv A'_{xz} N_2 + A'_{xx} Q_{22} A_{xi}, \\
\tilde{R}_{13} &\equiv -I/2, \\
\tilde{R}_{22} &\equiv R + A'_{xi} Q_{22} A_{xi} + 2A'_{xi} N_2.
\end{aligned}$$

Similarly, the law of motion for $\tilde{X}_t$ can then be written

$$\tilde{X}_{t+1} = \tilde{A}_{j_t j_{t+1}} \tilde{X}_t + \tilde{B}_{j_t j_{t+1}} \tilde{\imath}_t + \tilde{C}_{j_t j_{t+1}} \varepsilon_{t+1},$$

where

$$\tilde{A}_{jk} = \begin{bmatrix} A_{11k} + A_{12k} A_{xXj} & 0 \\ 0 & 0 \end{bmatrix},$$

$$\tilde{B}_{jk} = \begin{bmatrix} A_{12k} A_{xzj} & B_{1k} + A_{12k} A_{xij} & 0 \\ 0 & 0 & I \end{bmatrix}, \quad \tilde{C}_{jk} = \begin{bmatrix} C_k \\ 0 \end{bmatrix}.$$

C.1 Unobservable modes and forward-looking variables

The value function for the dual problem, $\tilde{V}(\tilde{X}_t, p_t)$, will be quadratic in $\tilde{X}_t$ for given p_t and can be written

$$\tilde{V}(\tilde{X}_t, p_t) \equiv \tilde{X}'_t \tilde{V}(p_t) \tilde{X}_t + w(p_t),$$

where

$$\tilde{V}(p_t) \equiv \sum\nolimits_j p_{jt} \hat{V}(p_t)_j, \qquad w(p_t) \equiv \sum\nolimits_j p_{jt} \hat{w}(p_t)_j.$$

Here, $\tilde{V}(p_t)$ and $\hat{V}(p_t)_j$ are symmetric $(n_X + n_x) \times (n_X + n_x)$ matrices and $w(p_t)$ and $\hat{w}(p_t)_j$ are scalars that are functions of p_t. (Thus, we simplify the notation and let $\tilde{V}(p_t)$ and $\hat{V}(p_t)_j$ $(j \in N_j)$ denote the matrices $\tilde{V}_{\tilde{X}\tilde{X}}(p_t)$ and $\hat{V}_{XX}(p_t, j_t)$ in section 2.3.2.) They will satisfy the Bellman equation for the dual saddlepoint problem,

$$\tilde{X}'_t \tilde{V}(p_t) \tilde{X}_t + w(p_t) = \max_{\gamma_t} \min_{z_t, i_t} \sum_j p_{jt} \left\{ \begin{array}{l} \tilde{X}'_t \tilde{Q}_j \tilde{X}_t + 2\tilde{X}'_t \tilde{N}_j \tilde{\imath}_t + \tilde{\imath}'_t \tilde{R}_j \tilde{\imath}_t \\ + \delta \sum_k P_{jk} [\tilde{X}'_{t+1,jk} \hat{V}(P' p_t)_k \tilde{X}_{t+1,jk} + \hat{w}(P' p_t)_k] \end{array} \right\},$$

where

$$\tilde{X}_{t+1,jk} \equiv \tilde{A}_{jk} \tilde{X}_t + \tilde{B}_{jk} \tilde{\imath}_t + \tilde{C}_{jk} \varepsilon_{t+1}.$$

The first-order condition with respect to $\tilde{\imath}_t$ is thus

$$\sum_j p_{jt} \left[\tilde{X}'_t \tilde{N}_j + \tilde{\imath}'_t \tilde{R}_j + \delta \sum_k P_{jk} (\tilde{X}'_t \tilde{A}'_{jk} + \tilde{\imath}'_t \tilde{B}'_{jk}) \hat{V}(P' p_t)_k \tilde{B}_{jk} \right] = 0,$$

which can be rewritten as

$$\sum_j p_{jt}\left[\tilde{N}_j'\tilde{X}_t + \tilde{R}_j\tilde{\imath}_t + \delta\sum_k P_{jk}\tilde{B}_{jk}'\hat{V}(P'p_t)_k(\tilde{A}_{jk}\tilde{X}_t + \tilde{B}_{jk}\tilde{\imath}_t)\right] = 0.$$

It is then apparent that the first-order conditions can be written compactly as

$$J(p_t)\tilde{\imath}_t + K(p_t)\tilde{X}_t = 0, \tag{C.2}$$

where

$$J(p_t) \equiv \sum_j p_{jt}\left[\tilde{R}_j + \delta\sum_k P_{jk}\tilde{B}_{jk}'\hat{V}(P'p_t)_k\tilde{B}_{jk}\right],$$

$$K(p_t) \equiv \sum_j p_{jt}\left[\tilde{N}_j' + \delta\sum_k P_{jk}\tilde{B}_{jk}'\hat{V}(P'p_t)_k\tilde{A}_{jk}\right]$$

This leads to the optimal policy function,

$$\tilde{\imath}_t = \tilde{F}(p_t)\tilde{X}_t,$$

where

$$\tilde{F}(p_t) \equiv -J(p_t)^{-1}K(p_t).$$

Furthermore, the value-function matrix $\tilde{V}(p_t)$ for the dual saddlepoint problem satisfies

$$\tilde{X}_t'\tilde{V}(p_t)\tilde{X}_t \equiv \sum_j p_{jt}\left\{\begin{array}{l}\tilde{X}_t'\tilde{Q}_j\tilde{X}_t + 2\tilde{X}_t'\tilde{N}_j\tilde{F}(p_t)\tilde{X}_t + \tilde{X}_t'\tilde{F}(p_t)'\tilde{R}_j\tilde{F}(p_t)\tilde{X}_t \\ +\,\delta\sum_k P_{jk}\tilde{X}_t'[\tilde{A}_{jk}' + \tilde{F}(p_t)'\tilde{B}_{jk}']\hat{V}(P'p_t)_k[\tilde{A}_{jk} + \hat{B}_{jk}\tilde{F}(p_t)]\tilde{X}_t\end{array}\right\}.$$

This implies the following Riccati equations for the matrix functions $\hat{V}(p_t)_j$,

$$\begin{aligned}\hat{V}(p_t)_j = {}& \tilde{Q}_j + \tilde{N}_j\tilde{F}(p_t) + \tilde{F}(p_t)'\tilde{N}_j' + \tilde{F}(p_t)'\tilde{R}_j\tilde{F}(p_t) \\ &+\, \delta\sum_k P_{jk}[\tilde{A}_{jk}' + \tilde{F}(p_t)'\tilde{B}_{jk}']\hat{V}(P'p_t)_k[\tilde{A}_{jk} + \tilde{B}_{jk}\tilde{F}(p_t)].\end{aligned}$$

The scalar functions $\hat{w}(p_t)_j$ will satisfy the equations

$$\hat{w}(p_t)_j = \delta\sum_k P_{jk}[\mathrm{tr}(\hat{V}(P'p_t)_k\tilde{C}_{jk}\tilde{C}_{jk}') + \hat{w}(P'p_t)_k].$$

The value function for the primal problem satisfies

$$\tilde{X}_t'V(p_t)\tilde{X}_t + w(p_t) \equiv \tilde{X}_t'\tilde{V}(p_t)\tilde{X}_t + w(p_t) - \Xi_{t-1}'\frac{1}{\delta}\sum\nolimits_j p_{jt}H_jF_x(p_t)_j\tilde{X}_t,$$

where we use that by (C.1) the equilibrium solution for x_t can be written

$$x_t = F_x(p_t)_j\tilde{X}_t.$$

We may also define the conditional value function for the primal problem as

$$\tilde{X}_t'V(p_t)_j\tilde{X}_t + w(p_t)_j \equiv \tilde{X}_t'\hat{V}(p_t)_j\tilde{X}_t + w(p_t)_j - \Xi_{t-1}'\frac{1}{\delta}H_jF_x(p_t)_j\tilde{X}_t \qquad (j \in N_j).$$

C.2 An algorithm for unobservable modes with forward-looking variables

Consider an algorithm for determining $\tilde{F}(p_t)$, $\tilde{V}(p_t)$, $w(p_t)$, $\hat{V}(p_t)_j$ and $\hat{w}(p_t)_j$ for a given distribution of the modes in period t, p_t. In order to get a starting point for the iteration, we assume that the modes become observable $T+1$ periods ahead, that is, in period $t+T+1$. Hence, from that period on, the relevant solution is given by the matrices $\tilde{F}_j$ and $\tilde{V}_j$ and scalars w_j for $j \in N_j$, where $\tilde{F}_j$ is the optimal policy function, $\tilde{V}_j$ is the value-function matrix, and w_j is the scalar in the value function for the dual saddlepoint problem with observable modes determined by the algorithm in appendix B.

We consider these matrices $\tilde{V}_j$ and scalars w_j and the horizon T as known, and we will consider an iteration for $\tau = T, T-1, ..., 0$ that determines $\tilde{F}(p_t)$, $\tilde{V}(p_t)$, and $w(p_t)$ as a function of T. The horizon T will then be increased until $\tilde{F}(p_t)$, $\tilde{V}(p_t)$, and $w(p_t)$ have converged.

Let $p_{t+\tau,t}$ for $\tau = 0, ..., T$ and given p_t be determined by the prediction equation,

$$p_{t+\tau,t} = (P')^\tau p_t,$$

and let $\hat{V}_k^{T+1} = \tilde{V}_k$ and $\hat{w}_k^{T+1} = w_k$ $(k \in N_j)$. Then, for $\tau = T, T-1, ..., 0$, let the mode-dependent matrices $\hat{V}_j^\tau$ and the mode-*in*dependent matrices $\tilde{V}^\tau$ and F^τ be determined recursively by

$$J^\tau \equiv \sum_j p_{j,t+\tau,t} \left[\tilde{R}_j + \delta \sum_k P_{jk} \tilde{B}'_{jk} \hat{V}_k^{\tau+1} \tilde{B}_{jk} \right], \tag{C.3}$$

$$K^\tau \equiv \sum_j p_{j,t+\tau,t} \left[\tilde{N}'_j + \delta \sum_k P_{jk} \tilde{B}'_{jk} \hat{V}_k^{\tau+1} \tilde{A}_{jk} \right], \tag{C.4}$$

$$\tilde{F}^\tau = -(J^\tau)^{-1} K^\tau, \tag{C.5}$$

$$\begin{aligned} \hat{V}_j^\tau = {} & \tilde{Q}_j + \tilde{N}_j \tilde{F}^\tau + \tilde{F}^{\tau\prime} \tilde{N}'_j + \tilde{F}^{\tau\prime} \tilde{R}_j \tilde{F}^\tau \\ & + \delta \sum_k P_{jk} [\tilde{A}'_{jk} + \tilde{F}^{\tau\prime} \tilde{B}'_{jk}] \hat{V}_k^{\tau+1} [\tilde{A}_{jk} + \tilde{B}_k \tilde{F}^\tau)], \end{aligned} \tag{C.6}$$

$$\hat{w}_j^\tau = \delta \sum_k P_{jk} [\mathrm{tr}(\hat{V}_k^{\tau+1} \tilde{C}_{jk} \tilde{C}'_{jk}) + \hat{w}_k^{\tau+1}], \tag{C.7}$$

$$\tilde{V}^\tau = \sum_j p_{j,t+\tau,t} \hat{V}_j^\tau, \tag{C.8}$$

$$w_j^\tau = \sum_j p_{j,t+\tau,t} \hat{w}_j^\tau. \tag{C.9}$$

This iteration will give $\tilde{F}^0$, $\tilde{V}^0$ and w^0 as functions of T. We let T increase until $\tilde{F}^0$ and $\tilde{V}^0$ have converged. Then, $\tilde{F}(p_t) = \tilde{F}^0$, $\tilde{V}(p_t) = \tilde{V}^0$, and $w(p_t) = w^0$. The value-function matrix $V(p_t)$ (denoted $V_{\tilde{X}\tilde{X}}(p_t)$ in section 2.3.2) for the primal problem will be given by

$$V(p_t) \equiv \tilde{V}(p_t) - \begin{bmatrix} 0 & \frac{1}{2}\Gamma_X(p_t)' \\ \frac{1}{2}\Gamma_X(p_t) & \frac{1}{2}[\Gamma_\Xi(p_t) + \Gamma_\Xi(p_t)'] \end{bmatrix},$$

where the matrix function

$$[\Gamma_X(p_t) \ \ \Gamma_\Xi(p_t)] \equiv \frac{1}{\delta}\sum\nolimits_j p_{jt}H_j[F_{xX}(p_t)_j \ \ F_{x\Xi}(p_t)_j]$$

is partitioned conformably with X_t and Ξ_{t-1}. The conditional value function matrix $V(p_t)_j$ for the primal problem will be given by

$$V(p_t)_j \equiv \hat{V}(p_t)_j - \begin{bmatrix} 0 & \frac{1}{2}\Gamma_X(p_t)_j' \\ \frac{1}{2}\Gamma_X(p_t)_j & \frac{1}{2}[\Gamma_\Xi(p_t)_j + \Gamma_\Xi(p_t)_j'] \end{bmatrix} \qquad (j \in N_j),$$

where $\hat{V}(p_t)_j = \hat{V}_j^0$ and the matrix function

$$[\Gamma_X(p_t)_j \ \ \Gamma_\Xi(p_t)_j] \equiv \frac{1}{\delta}H_j[F_{xX}(p_t)_j \ \ F_{x\Xi}(p_t)_j]$$

is partitioned conformably with X_t and Ξ_{t-1}.

C.3 An algorithm for unobservable modes without forward-looking variables

When there are no forward-looking variables, the primal loss function can be written

$$\tilde{L}_t = \begin{bmatrix} X_t \\ i_t \end{bmatrix}' \begin{bmatrix} Q_j & N_j \\ N_j' & R_j \end{bmatrix} \begin{bmatrix} X_t \\ i_t \end{bmatrix},$$

with $\tilde{X}_t = X_t$ and $\tilde{\imath}_t = i_t$. There is no need for the dual optimization problem, and the algorithm simply applies (C.3)-(C.9) for the determination of $F^\tau = \tilde{F}^\tau$ and $V^\tau = \tilde{V}^\tau$ with the matrices $\tilde{A}_{jk} = A_k$, $\tilde{B}_{jk} = B_k$, $\tilde{C}_{jk} = C_k$, $\tilde{Q}_j = Q_j$, $\tilde{N}_j = N_j$, and $\tilde{R}_j = R_j$, in which case the optimal policy function on period t is $i_t = F(p_t)X_t$ with $F(p_t) = F^0$ and the value function is $X_t'V(p_t)X_t + w(p_t)$ with $V(p_t) = V^0 = \tilde{V}^0$ and $w(p_t) = w^0$.

D A unit discount factor

The expected discounted losses (2.4) and (2.13) are normally bounded only for $\delta < 1$. More precisely, w_j $(j \in N_j)$ in (2.21) is normally bounded only for $\delta < 1$. The case $\delta = 1$ can be handled by scaling the intertemporal loss function by $1-\delta$ for $\delta < 1$ and then consider the limit when $\delta \to 1$, as mentioned in footnote 7. That is, we can replace the intertemporal loss function in (2.4) and (2.13) by $\mathrm{E}_t(1-\delta)\sum_{\tau=0}^{\infty}\delta^\tau L_{t+\tau}$ and $\mathrm{E}_t(1-\delta)\sum_{\tau=0}^{\infty}\delta^\tau \tilde{L}_{t+\tau}$, respectively. In particular, we can write (2.21) as

$$(1-\delta)\tilde{X}_t'V_{j_t}\tilde{X}_t + \delta w_j \equiv \min \mathrm{E}_t(1-\delta)\sum_{\tau=0}^{\infty}\delta^\tau L_{t+\tau}. \tag{D.1}$$

Then, V_j $(j \in N_j)$ is still determined as before, whereas w_j now satisfies

$$w_j(\delta) = \sum_k P_{jk}\{(1-\delta)\mathrm{tr}[V_k(\delta)\tilde{C}_k\tilde{C}_k'] + \delta w_k(\delta)\} \qquad (j \in N_j), \tag{D.2}$$

where our notation emphasizes that w_j and V_j depend on δ.

From (D.1), we see that

$$\lim_{\delta \to 1-} \min \mathrm{E}_t (1-\delta) \sum_{\tau=0}^{\infty} \delta^\tau L_{t+\tau} = w_j(1) \qquad (j \in N_j).$$

Furthermore, from (D.2), we see that

$$w_j(1) = \sum_k P_{jk} w_k(1) \qquad (j \in N_j),$$

so the vector $(w_1(1), ..., w_{n_j}(1))'$ is an eigenvector for the eigenvalue 1 of the transition matrix P. By our assumptions on the Markov chain in footnote 5, the Markov chain is fully regular, so the only such eigenvector is $(1, ..., 1)$ (and scalar multiples thereof) (Gantmacher [18]). Therefore, $w_j(1)$ is independent of j:

$$w_j(1) = w \qquad (j \in N_j)$$

for some scalar w.

For $\delta < 1$, we multiply (D.2) by $\bar{p}_j$ and sum over j. This gives

$$\sum_j \bar{p}_j w_j(\delta) = \sum_j \sum_k \bar{p}_j P_{jk} \{(1-\delta) \mathrm{tr}[V_k(\delta) \tilde{C}_k \tilde{C}_k'] + \delta w_k\} = (1-\delta) \sum_k \bar{p}_k \mathrm{tr}[V_k(\delta) \tilde{C}_k \tilde{C}_k'] + \delta \sum_k \bar{p}_k w_k(\delta).$$

Letting $\bar{w}(\delta) \equiv \sum_j \bar{p}_j w_j(\delta)$, we see that

$$\bar{w}(\delta) = \sum_k \bar{p}_k \mathrm{tr}[V_k(\delta) \tilde{C}_k \tilde{C}_k'].$$

We conclude that in the limit, when $\delta \to 1$, the expected minimum loss is given by

$$w_j(1) = \bar{w}(1) = \sum_k \bar{p}_k \mathrm{tr}[V_k(1) \tilde{C}_k \tilde{C}_k'] \qquad (j \in N_j)$$

and is independent of $\tilde{X}_t$ and j_t. Intuitively, for $\delta \to 1$, current losses become insignificant relative to expected losses far into the future, and then the stationary distribution $\bar{p}$ applies. Therefore, the expected discounted loss becomes independent of both the current predetermined variables and the current mode, even though the optimal policy function depends on the current mode (when the modes are observable) or the distribution of the current modes (when the modes are unobservable).

E Mean square stability

Costa, Fragoso, and Marques [11, chapter 3] (CFM) provide a discussion of stability for MJLQ systems. An appropriate concept of stability for our purpose is mean square stability, which is defined as follows:

Consider the system

$$X_{t+1} = \Gamma_{\theta_t} X_t,$$

for $t = 0, 1, ...,$ where $X_t \in R^{n_X}$, $\theta_t \in \Theta \equiv \{1, ..., N\}$ is a Markov process with transition probabilities $\mathcal{P}_{jk} = \Pr\{\theta_{t+1} = k \,|\, \theta_t = j\}$ $(j, k \in \Theta)$, transition matrix $\mathcal{P} = [\mathcal{P}_{jk}]$, and Γ_θ is an $n_X \times n_X$ matrix that depends on $\theta \in \Theta$, and $X_0 \in R^{n_X}$ and $\theta_0 \in \Theta$ are given. The system is *mean square stable* (MSS) if, for any initial $X_0 \in R^{n_X}$ and $\theta_0 \in \Theta$, there exist a vector $\mu \in R^{n_X}$ and an $n_X \times n_X$ matrix Q independent of X_0 and θ_0 such that $||\mathrm{E}[X_t] - \mu|| \to 0$ and $||\mathrm{E}[X_t X_t'] - Q|| \to 0$ when $t \to \infty$.

CFM [11, theorem 3.9] provide six equivalent necessary and sufficient conditions for mean square stability. The following necessary and sufficient condition is appropriate for our purpose:

Define the matrices $\mathcal{C}$ and $\mathcal{N}$ by

$$\mathcal{C} \equiv \mathcal{P}' \otimes I_{n_X^2},$$

$$\mathcal{N} \equiv \mathrm{diag}(\Gamma_\theta \otimes \Gamma_\theta) \equiv \begin{bmatrix} \Gamma_1 \otimes \Gamma_1 & 0 & \cdots & 0 \\ 0 & \Gamma_2 \otimes \Gamma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Gamma_N \otimes \Gamma_N \end{bmatrix}.$$

The system above is MSS if and only if the spectral radius (the supremum of the modulus of the eigenvalues) of the matrix $\mathcal{CN}$ is less than unity.

Applying CFM's definition of and conditions for mean square stability requires a simple redefinition of the modes in our framework. Start from the system

$$X_{t+1} = M_{j_t j_{t+1}} X_t,$$

where $t = 0, 1, ...,$ $X_t \in R^{n_X}$, $j_t \in \{1, ..., n_j\}$, $P = [P_{jk}]$, $P_{jk} = \Pr\{j_{t+1} = k \,|\, j_t = j\}$, and X_0 and j_0 are given. This system differs from CFM's system in that the matrix $M_{j_t j_{t+1}}$ depends on the realization of the modes in both period t and period $t+1$.

Define the new composite mode $\theta_t \equiv (j_t, j_{t+1})$, which can take $N = n_j^2$ values, and consider a Markov chain for θ_t with transition probabilities $\mathcal{P}_{\theta\kappa} \equiv \Pr\{\theta_{t+1} = \kappa \equiv (k, l) \,|\, \theta_t = \theta \equiv (j, k)\}$. We note that the transition probability from $\theta_t = (j, k)$ to $\theta_{t+1} = (k, l)$ does not depend on j but only on k and l. Furthermore, it is simply P_{kl}, so

$$\mathcal{P}_{(j,k),(k,l)} = P_{kl} \qquad (j, k, l = 1, ..., n_j).$$

Thus, we can consider the new system

$$X_{t+1} = M_{\theta_t} X_t,$$

where θ_t is a Markov chain that can take n_j^2 different values and has a transition matrix $\mathcal{P}$ with the transition probabilities $\mathcal{P}_{\theta_t \theta_{t+1}}$ defined above. Then the results of CFM on MSS apply directly, and we only need to define Γ_θ, $\mathcal{P}$, $\mathcal{C}$, and $\mathcal{N}$ using the n_j^2-mode composite Markov chain for $\theta_t \equiv (j_t, j_{t+1})$ instead of just the n_j-mode chain for j_t.

F Details of the estimation

Here we lay out the details of the priors we use in our Bayesian estimation.

For the RS model in section 4.1, we base our prior for the MJLQ case on our OLS estimates. The priors are identical across modes. In particular, the priors for the vectors of coefficients α_i and β_i are, except for α_3 and β_3, each multivariate normal distributions, with mean given by the OLS point estimates and a covariance matrix given by the covariance matrix of the estimates scaled up by a factor of 4. The coefficient α_3 is restricted to be positive and the coefficient β_3 is restricted to be negative, and the priors for α_3 and $-\beta_3$ are lognormal with means that match the constant-coefficient case and variances that are four times the estimated variance from the constant-coefficient case. For the parameters of the transition matrix P of the Markov chain, we take independent beta distributions (subject to the constraint that the rows sum to one). We let the diagonal elements have mean 0.9 and standard deviation 0.08, while the off-diagonals have means 0.05 and standard deviations 0.05. For the variances of the shocks, we assume an inverse gamma prior distribution with two degrees of freedom.

For the Lindé model in section 4.2, we take independent priors for the different structural coefficients, again with the priors being identical across modes. For the coefficients ω_f and β_f, we assume a beta distribution with mean 0.5 and standard deviation 0.25. The other structural coefficients have normal distributions, with $\gamma \sim N(0.1, 0.05^2)$, $\beta_r \sim N(0.15, 0.075^2)$, $\beta_y \sim N(1.5, 0.5^2)$, $\rho_1 \sim N(0.9, 0.2^2)$, $\rho_2 \sim N(0.2, 0.2^2)$, $\gamma_\pi \sim N(1.5, 0.5^2)$, and $\gamma_y \sim N(0.5, 0.5^2)$. Again for the variances of the shocks, we assume an inverse gamma prior distribution with two degrees of freedom. The prior over the Markov chain transition matrix is the same as in the RS model.

G An algorithm for an arbitrary instrument rule

Consider the case when the time-invariant instrument rule can be written

$$i_t = F_{Xj_t} X_t + F_{xj_t} x_t \qquad (j_t = 1, ..., n_j). \tag{G.1}$$

If there is a unique solution associated with a specified instrument rule, it will determine the forward-looking variables as a linear function of the predetermined variables,

$$x_t = G_{j_t} X_t.$$

Given a quadratic intertemporal loss function, this will also determine a value of the loss function of the form

$$X_t' V_{j_t} X_t + w_{j_t}.$$

In order to specify an algorithm for finding G_j, V_j, and w_j, suppose the instrument rule can be written as (G.1). Consider period $t+1$, and assume that $G_{j_{t+1}}^{(t+1)}$ in

$$x_{t+1} = G_{j_{t+1}}^{(t+1)} X_{t+1} \qquad (j_{t+1} = 1, ..., n_j),$$

is known in period t. This will imply

$$\begin{aligned}
\mathrm{E}_t H_{j_{t+1}} x_{t+1} &= \mathrm{E}_t H_{j_{t+1}} G_{j_{t+1}}^{(t+1)} X_{t+1} \\
&= \sum_k P_{jk} H_k G_k^{(t+1)} [(A_{11k} + B_{1k} F_{Xj}) X_t + (A_{12k} + B_{1k} F_{xj}) x_t] \\
&= (A_{21j} + B_{2j} F_{Xj}) X_t + (A_{22j} + B_{2j} F_{xj}) x_t.
\end{aligned}$$

We can then solve for x_t in period t,

$$x_t = G_j^{(t)} X_t,$$

where

$$\begin{aligned}
G_j^{(t)} &\equiv \left[A_{22j} + B_{2j} F_{xj} - \sum_k P_{jk} H_k G_k^{(t+1)} (A_{12k} + B_{1k} F_{xj}) \right]^{-1} \\
&\quad \cdot \left[\sum_k P_{jk} H_k G_k^{(t+1)} (A_{11k} + B_{1k} F_{Xj}) - (A_{21j} + B_{2j} F_{Xj}) \right].
\end{aligned}$$

It follows that, starting with a guess G_j^0, the iteration for $l = 0, 1, ...,$ according to

$$\begin{aligned}
G_j^{l+1} &= \left[A_{22j} + B_{2j} F_{xj} - \sum_k P_{jk} H_k G_k^l (A_{12k} + B_{1k} F_{xj}) \right]^{-1} \\
&\quad \cdot \left[\sum_k P_{jk} H_k G_k^l (A_{11k} + B_{1k} F_{Xj}) - (A_{21j} + B_{2j} F_{Xj}) \right],
\end{aligned}$$

will hopefully make G_j^l converge to the correct G_j,

$$x_t = G_j X_t. \tag{G.2}$$

This then implies

$$X_{t+1} = M_{jk} X_t + C_k \varepsilon_{t+1},$$

where

$$M_{jk} \equiv A_{11k} + A_{12k} G_j + B_{1k} (F_{Xj} + F_{xj} G_j).$$

Clearly, $G \equiv \{G_j\}$ and $M \equiv \{M_{jk}\}$ will be functions of $F \equiv \{(F_{Xj}, F_{xj})\}$.

Let the period loss function be

$$L_t = \begin{bmatrix} \begin{bmatrix} X_t \\ x_t \end{bmatrix} \\ i_t \end{bmatrix}' \begin{bmatrix} Q & N \\ N' & R \end{bmatrix} \begin{bmatrix} \begin{bmatrix} X_t \\ x_t \end{bmatrix} \\ i_t \end{bmatrix}. \tag{G.3}$$

Given (G.1), (G.2), and (G.3), we can define the matrix

$$\bar{W}_j \equiv \begin{bmatrix} \begin{bmatrix} I \\ G_j \end{bmatrix} \\ F_{Xj} + F_{xj} G_j \end{bmatrix}' \begin{bmatrix} Q & N \\ N' & R \end{bmatrix} \begin{bmatrix} \begin{bmatrix} I \\ G_j \end{bmatrix} \\ F_{Xj} + F_{xj} G_j \end{bmatrix},$$

in which case the period loss satisfies

$$L_t = X_t' \bar{W}_j X_t.$$

It follows that the value function corresponding to the intertemporal loss function

$$\mathrm{E}_t \sum_{\tau=0}^{\infty} \delta^\tau L_{t+\tau}$$

will satisfy

$$X_t' V_j X_t + w_j = X_t' \bar{W}_j X_t + \delta \sum_k P_{jk} [X_t' M_{jk}' V_k M_{jk} X_t + \mathrm{tr}(V_k C_k C_k') + w_k].$$

Hence, the matrix V_j will satisfy the Lyapunov equation

$$V_j = \bar{W}_j + \delta \sum_k P_{jk} M_{jk}' V_k M_{jk} \qquad (j \in N_j),$$

and w_j will satisfy

$$w_j = \delta \sum_k P_{jk} [\mathrm{tr}(V_k C_k C_k') + w_k] \qquad (j \in N_j).$$

Note that we can, for each j, define

$$\begin{aligned} \hat{W}_j &\equiv \bar{W}_j + \delta \sum_{k \neq j} P_{jk} M_{jk}' V_k M_{jk} \\ \hat{M}_{jj} &= \sqrt{\delta P_{jj}} M_{jj}, \end{aligned}$$

and then solve the more standard Lyapunov equation

$$V_j = \hat{W}_j + \hat{M}_{jj}' V_j \hat{M}_{jj} \qquad (j \in N_j).$$

Clearly, $V \equiv \{V_j\}$ and $w \equiv \{w_j\}$ will be functions of F and δ.

Let $\bar{p}_j$ $(j \in N_j)$ denote the stationary distribution of the states, and let $\bar{V} \equiv \sum_j \bar{p}_j V_j$ and $\bar{w} \equiv \sum_j \bar{p}_j w_j$ denote the unconditional means of V_j and w_j. We note that

$$\bar{w} = \frac{\delta}{1-\delta} \sum_k \bar{p}_k \mathrm{tr}(V_k C_k C_k').$$

Suppose that the intertemporal loss function is $1-\delta$ times the one above,

$$\mathrm{E}_t \sum_{\tau=0}^{\infty} (1-\delta) \delta^\tau L_{t+\tau},$$

and suppose that we consider the limit when $\delta \to 1$,

$$\lim_{\delta \to 1} \mathrm{E}_t \sum_{\tau=0}^{\infty} (1-\delta) \delta^\tau L_{t+\tau} = \mathrm{E}[L_t].$$

In that case, the intertemporal loss function is just the unconditional mean of the period loss function, $\mathrm{E}[L_t]$. Furthermore, the unconditional mean of $1-\delta$ times the value function above will be

$$(1-\delta)\{\mathrm{E}[X_t' V_{j_t} X_t] + \bar{w}\} = (1-\delta)\mathrm{E}[X_t' V_t X_t] + \delta \sum_k \bar{p}_k \mathrm{tr}(V_k C_k C_k').$$

We see that, when $\delta \to 1$, the first term on the right side goes to zero, and we conclude that, in the limit,

$$\mathrm{E}[L_t] = \sum_k \bar{p}_k \mathrm{tr}[V_k(F,1) C_k C_k'],$$

where we also explicitly note that V_k depends on F and δ.

Suppose the instrument rule is restricted to a given class $\mathcal{F}$ of instrument rules

$$F \in \mathcal{F}.$$

The optimal instrument rule in this class, $\hat{F}$, can now be defined as

$$\hat{F} \equiv \arg\min_{F \in \mathcal{F}} \sum_k \bar{p}_k \mathrm{tr}[V_k(F,1) C_k C_k'].$$

It will obviously depend on $C_k C_k'$, the covariance matrix of the shock $C_k \varepsilon_{t+1}$. Hence, certainty equivalence does generally not hold for optimal restricted instrument rules.

References

[1] Anderson, Evan W., Lars Peter Hansen, Ellen R. McGrattan, and Thomas J. Sargent (1995), "Mechanics of Forming and Estimating Dynamic Linear Economies," working paper, homepages.nyu.edu/∼ts43/.

[2] Aoki, Masanao (1967), *Optimization of Stochastic Systems*, Academic Press, New York.

[3] Backus, David, and John Driffill (1986), "The Consistency of Optimal Policy in Stochastic Rational Expectations Models," CEPR Discussion Paper No. 124.

[4] Blake, Andrew P., and Fabrizio Zampolli (2006), "Time Consistent Policy in Markov Switching Models," Working Paper No. 298, Bank of England.

[5] Blinder, Alan S., and Ricardo Reis (2005), "Understanding the Greenspan Standard," in *The Greenspan Era: Lessons for the Future*, a symposium sponsored by the Federal Reserve Bank of Kansas City, forthcoming, www.federalreserve.gov.

[6] Brainard, W. (1967), "Uncertainty and the Effectiveness of Policy," *American Economic Review* 57, 411–425.

[7] Brash, Donald T. (2000), "Making Monetary Policy: A Look Behind the Curtains," speech in Christchurch, January 26, 2000, Reserve Bank of New Zealand, www.rbnz.govt.nz.

[8] Chow, Gregory C. (1973) "Effect of Uncertainty on Optimal Control Policies," *International Economic Review* 14, 632–45.

[9] Cogley, Timothy, Riccardo Colacito and Thomas J. Sargent (2005) "The Benefits from U.S. Monetary Policy Experimentation in the Days of Samuelson and Solow and Lucas," working paper, NYU.

[10] Costa, Oswaldo L.V., and Marcelo D. Fragoso (1995), "Discrete-Time LQ-Optimal Control Problems for Infinite Markov Jump Parameter Systems," *IEEE Transactions on Automatic Control* 40, 2076–2088.

[11] Costa, Oswaldo L.V., Marecelo D. Fragoso, and Ricardo P. Marques (2005), *Discrete-Time Markov Jump Linear Systems*, Springer, London.

[12] Currie, David, and Paul Levine (1993), *Rules, Reputation and Macroeconomic Policy Coordination*, Cambridge University Press, Cambridge.

[13] Davig, Troy and Eric Leeper (2005), "Generalizing the Taylor Principle," NBER Working Paper 11874.

[14] do Val, João B.R., and Tamer Başar (1999), "Receding Horizon Control of Jump Linear Systems and a Macroeconomic Policy Problem," *Journal of Economic Dynamics and Control* 23, 1099–1131.

[15] do Val, João B.R, José C. Geromel, and Oswaldo L.V. Costa (1998), "Uncoupled Riccati Iterations for the Linear Quadratic Control Problem of Discrete-Time Markov Jump Linear Systems," *IEEE Transactions on Automatic Control* 43, 1727–1733.

[16] Farmer, Roger E. A., Daniel F. Waggoner, and Tao Zha (2006) "Understanding the New-Keynesian Model when Monetary Policy Switches Regimes," working paper.

[17] Feldstein, Martin (2004), "Innovations and Issues in Monetary Policy: Panel Discussion," *American Economic Review* 94 (May) 41–43.

[18] Gantmacher, Felix R. (1959), *The Theory of Matrices*, Volume II, Chelsea Publishing Company, New York.

[19] Greenspan, Alan (2004), "Risk and Uncertainty in Monetary Policy," *American Economic Review* 94 (May) 33–40.

[20] Greenspan, Alan (2005), "Reflections on Central Banking," in *The Greenspan Era: Lessons for the Future*, a symposium sponsored by the Federal Reserve Bank of Kansas City, forthcoming, www.federalreserve.gov.

[21] Hamilton, James D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica* 57, 357-384.

[22] Hamilton, James D. (1994), *Time Series Analysis*, Princeton University Press.

[23] Kahn, James A. and Robert W. Rich (2004) "Tracking the New Economy: Using Growth Theory to Detect Changes in Trend Productivity," working paper, Federal Reserve Bank of New York.

[24] Karlin, Samuel, and Howard M. Taylor (1975), *A First Course in Stochastic Processes*, 2nd edition, Academic Press, San Diego, CA.

[25] Kim, Chang-Jin and Charles R. Nelson (1999), *State-Space Models with Regime Switching*, MIT Press, Cambridge, MA.

[26] Levin, Andrew T., Alexei Onatski, John C. Williams, and Noah Williams (2005), "Monetary Policy Under Uncertainty in Micro-Founded Macroeconometric Models," working paper for the 2005 NBER Macroeconomics Annual.

[27] Lindé, Jesper (2005), "Estimating New-Keynesian Phillips Curves: A Full Information Maximum Likelihood Approach," *Journal of Monetary Economics* 52, 1135–1149.

[28] Ljungqvist, Lars, and Thomas J. Sargent (2005), *Recursive Macroeconomic Theory*, 2nd edition, MIT Press, Cambridge, MA.

[29] Marcet, Albert, and Ramon Marimon (1998), "Recursive Contracts," working paper, www.econ.upf.edu.

[30] McConnell, Margaret M. and Gabriel Perez-Quiros (2000) "Output Fluctuations in the United States: What Has Changed since the Early 1980's?", *American Economic Review* 90, 1464–76.

[31] Onatski, Alexei, and Noah Williams (2003), "Modeling Model Uncertainty," *Journal of the European Economic Association* 1, 1087–1022.

[32] Oudiz, Gilles, and Jeffrey Sachs (1985), "International Policy Coordination in Dynamic Macroeconomic Models," in William H. Buiter and Richard C. Marston, eds., *International Economic Policy Coordination*, Cambridge University Press, Cambridge.

[33] Rudebusch, Glenn D., and Lars E.O. Svensson (1999), "Policy Rules for Inflation Targeting," in John B. Taylor (ed.), *Monetary Policy Rules*, University of Chicago Press.

[34] Sims, Christopher A. (2002), "The Role of Models and Probabilities in the Monetary Policy Process," *Brookings Papers on Economic Activity* 2:2002, 1–62.

[35] Söderlind, Paul (1999), "Solution and Estimation of RE Macromodels with Optimal Policy," *European Economic Review* 43, 813–823.

[36] Stock, James H. and Mark W. Watson (2002) "Has the Business Cycle Changed and Why?" in Mark Gertler and Ken Rogoff, eds., *NBER Macroeconomics Annual 2002*, MIT Press, Cambridge.

[37] Svensson, Lars E.O. (2001a), *Independent Review of the Operation of Monetary Policy in New Zealand: Report to the Minister of Finance*, www.princeton.edu/svensson.

[38] Svensson, Lars E.O. (2001b), "Price Stability as a Target for Monetary Policy: Defining and Maintaining Price Stability," in Deutsche Bundesbank, ed., *The Monetary Transmission Process: Recent Developments and Lessons for Europe, Palgrave*, New York, 60–102, www.princeton.edu/svensson.

[39] Svensson, Lars E.O. (2005), "Monetary Policy with Judgment: Forecast Targeting," *International Journal of Central Banking* 1, 1–54, www.ijcb.org.

[40] Svensson, Lars E.O. (2006), "Optimization under Commitment and Discretion, the Recursive Saddlepoint Method, and Targeting Rules and Instrument Rules: Lecture Notes," www.princeton.edu/svensson.

[41] Svensson, Lars E.O., and Robert J. Tetlow (2005), "Optimal Policy Projections," *International Journal of Central Banking*, forthoming, www.princeton.edu/svensson.

[42] Svenssson, Lars E.O., and Noah Williams (2006), "Bayesian and Adaptive Optimal Policy under Model Uncertainty," working paper, www.princeton.edu/svensson and www.princeton.edu/˜noahw.

[43] Svensson, Lars E.O., and Michael Woodford (2005), "Implementing Optimal Policy through Inflation-Forecast Targeting," in Bernanke, Ben S., and Michael Woodford, eds., *Inflation Targeting*, University of Chicago Press, www.princeton.edu/svensson.

[44] Woodford, Michael (2003), *Interest and Prices*: *Foundations of a Theory of Monetary Policy*, Princeton University Press.

[45] Zampolli, Fabrizio (2006), "Optimal Monetary Policy in a Regime-Switching Economy: The Response to Abrupt Shifts in Exchange-Rate Dynamics," Working Paper No. 297, Bank of England.