BOP709.tex

Bayesian and Adaptive Optimal Policy under Model Uncertainty*

Lars E.O. Svensson
Sveriges Riksbank and Princeton University
www.princeton.edu/svensson/

Noah Williams
Princeton University
www.princeton.edu/∼noahw/

First draft: September 2005
This version: September 2007

Abstract

We study the problem of a policymaker who seeks to set policy optimally in an economy where the true economic structure is unobserved, and he optimally learns from observations of the economy. This is a classic problem of learning and control, variants of which have been studied in the past, but seldom with forward-looking variables which are a key component of modern policy-relevant models. As in most Bayesian learning problems, the optimal policy typically includes an experimentation component reflecting the endogeneity of information. We develop algorithms to solve numerically for the Bayesian optimal policy (BOP). However, computing the BOP is only feasible in relatively small models, and thus we also consider a simpler specification we term adaptive optimal policy (AOP) which allows policymakers to update their beliefs but shortcuts the experimentation motive. In our setting, the AOP is significantly easier to compute, and in many cases provides a good approximation to the BOP. We provide some simple examples to illustrate the role of learning and experimentation in an MJLQ framework.

JEL Classification: E42, E52, E58
Keywords: Optimal monetary policy, learning, recursive saddlepoint method

*The authors thank James Bullard, Martin Ellison, Alexei Onatski, Tamas Papp, Rujikorn Pavasuthipaisit, Eric Schaling, Eric Swanson, Rick van der Ploeg, members of the Princeton Macro Lunch Workshop, and participants in the following conferences: Monetary Policy and Uncertainty in Oslo, June 2006, Macroeconomic Policy Design for Monetary Union in Amsterdam, June 2006, and the International Research Forum on Monetary Policy in Washington, DC, December 2006, for helpful comments and discussions. The views expressed in this paper are solely the responsibility of the authors and should not to be interpreted as reflecting the views of any other member of the Executive Board of Sveriges Riksbank. Financial support from the National Science Foundation is gratefully acknowledged.

1 Introduction

We study the problem of a policymaker (more concretely, a central bank), who seeks to set policy optimally in an economy where the true economic structure is unobserved and the policymaker optimally learns from his observations of the economy. This is a classic problem of learning and control with model uncertainty, variants of which have been studied in the past, but very little has been done with forward-looking variables, which are a key component of modern policy-relevant models. We gain some tractability by taking our model of the economy to be a so-called Markov jump-linear-quadratic (MJLQ) system, extended to include forward-looking variables. In this setup, model uncertainty takes the form of different "modes" or regimes that follow a Markov process. This setup can be adapted to handle many different forms of model uncertainty, but yet provides a relatively simple structure for analysis.

In a previous paper, Svensson and Williams [15], we studied optimal policy design in models of this class when policymakers can or cannot observe the current mode, but we abstracted from any learning and inference about the current mode. In this paper we focus on learning and inference in the more relevant situation, particularly for the model-uncertainty applications which interest us, in which the modes are not directly observable. Thus, decision makers must filter their observations to make inferences about the current mode. As in most Bayesian learning problems, the optimal policy thus typically includes an experimentation component reflecting the endogeneity of information. This class of problems has a long history in economics, and it is well-known that solutions are difficult to obtain. We develop algorithms to solve numerically for the optimal policy.[1] Due to the curse of dimensionality, the Bayesian optimal policy (BOP) is only feasible in relatively small models. Confronted with these difficulties, we also consider *adaptive* optimal policy (AOP).[2] In this case, the policymaker in each period does update the probability distribution of the current mode in a Bayesian way, but the optimal policy is computed each period under the assumption that the policymaker will not learn in the future from observations. In our MJLQ setting, the AOP is significantly easier to compute, and in many cases provides a good approximation to the

[1] In addition to the classic literature (on such problems as a monopolist learning its demand curve), Wieland [17]-[18] and Beck and Wieland [1] have recently examined Bayesian optimal policy and optimal experimentation in a context similar to ours but without forward-looking variables. Tesfaselassie, Schaling, and Eijffinger [16] examine passive and active learning in a simple model with a forward-looking element in the form of a long interest rate in the aggregate-demand equation. Ellison and Valla [7] and Cogley, Colacito, and Sargent [3] study situations like ours but where the expectational component is as in the Lucas-supply curve ($E_{t-1}\pi_t$, for example) rather than our forward-looking case ($E_t\pi_{t+1}$, for example). More closely related to our present paper, Ellison [6] analyzes active and passive learning in a New Keynesian model with uncertainty about the slope of the Phillips curve.

[2] What we call optimal policy under no learning, adaptive optimal policy, and Bayesian optimal policy has in the literature also been referred to as myopia, passive learning, and active learning, respectively.

BOP. Moreover, the AOP analysis is of some interest in its own right, as it is closely related to specifications of adaptive learning which have been widely studied in macroeconomics (see Evans and Honkapohja [8] for an overview). Further, the AOP specification rules out the experimentation which some may view as objectionable in a policy context.[3]

In this paper, we provide two simple examples, with and without forward-looking variables, to illustrate the role of learning and experimentation in an MJLQ framework and compare the policy functions and value functions under NL, AOP, and BOP. Of particular interest is how uncertainty affects policy, and how learning interacts with the optimal policy decisions. We also diagnose the aspects of the model which influence the size of experimentation motive, and thus drive the differences between the Bayesian and adaptive optimal policies.

MJLQ models have been widely studied in the control-theory literature for the special case when the model modes are observable and there are no forward-looking variables (see Costa, Fragoso, and Marques [4] (henceforth CFM) and the references therein).[4] More recently, Zampolli [19] has used such an MJLQ model to examine monetary policy under shifts between regimes with and without an asset-market bubble. Blake and Zampolli [2] provide an extension of the MJLQ model with observable modes to include forward-looking variables and present an algorithm for the solution of an equilibrium resulting from optimization under discretion. Svensson and Williams [15] provide a more general extension of the MJLQ framework with forward-looking variables and present algorithms for the solution of an equilibrium resulting from optimization under commitment in a timeless perspective as well as arbitrary time-varying or time-invariant policy rules, using the recursive saddlepoint method of Marcet and Marimon [11]. They also provide two concrete examples: an estimated backward-looking model (a three-mode variant of Rudebusch and Svensson [13]) and an estimated forward-looking model (a three-mode variant of Lindé [10]). Svensson and Williams [15] also extend the MJLQ framework to the more realistic case of unobservable modes, although without introducing learning and inference about the probability distribution of modes, which is our focus here.

The paper is organized as follows: Section 2 lays out the basic model an MJLQ system with forward-looking variables. Sections 3, 4, and 5 specify the optimal policy under no learning (NL), the adaptive optimal policy (AOP), and the Bayesian optimal policy (BOP). Section 6 provides some simple examples, and compares the value functions and policy functions for these three alternatives

[3] In addition, AOP is useful for technical reasons as it gives us a good starting point for our more intensive numerical calculations in the BOP case.

[4] do Val and Başar [5] provide an application of an adaptive-control MJLQ problem in economics.

and clarifies the benefits and costs of optimal experimentation.

2 The model

We consider a Markov Jump-Linear-Quadratic (MJLQ) model of an economy with forward-looking variables. The economy has a private sector and a policymaker. We let X_t denote an n_X-vector of predetermined variables in period t, x_t an n_x-vector of forward-looking variables, and i_t an n_i-vector of (policymaker) instruments (control variables).[5] We let model uncertainty be represented by n_j possible modes and let $j_t \in N_j \equiv \{1, 2, ..., n_j\}$ denote the mode in period t. The model of the economy can then be written

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

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

where ε_t is a multivariate normally distributed random i.i.d. n_ε-vector of shocks with mean zero and contemporaneous covariance matrix I_{n_ε}. The matrices A_{11j}, A_{12j}, ..., C_{2j} have the appropriate dimensions and depend on the mode j. As a structural model here is simply a collection of matrices, each mode can represent a different model of the economy. Thus, uncertainty about the prevailing mode *is* model uncertainty.[6]

Note that the matrices on the right side of (2.1) depend on the mode j_{t+1} in period $t+1$, whereas the matrices on the right side of (2.2) depend on the mode j_t in period t. Equation (2.1) then determines the predetermined variables in period $t+1$ as a function of the mode and shocks in period $t+1$ and the predetermined variables, forward-looking variables, and instruments in period t. Equation (2.2) determines the forward-looking variables in period t as a function of the mode and shocks in period t, the expectations in period t of next period's mode and forward-looking variables, and the predetermined variables and instruments in period t. The matrix A_{22j} is invertible for each $j \in N_j$.

The mode j_t follows a Markov process with the transition matrix $P \equiv [P_{jk}]$.[7] The shocks ε_t are mean zero and i.i.d. with density φ, and without loss of generality we assume that ε_t is independent j_t.[8] We also assume that $C_{1j}\varepsilon_t$ and $C_{2k}\varepsilon_t$ are independent for all $j, k \in N_j$. These shocks, along

[5] The first component of X_t may be unity, in order to allow for mode-dependent intercepts in the model equations.

[6] See also Svensson and Williams [15], where we show how many different types of uncertainty can be mapped into our MJLQ framework.

[7] Obvious special cases are $P = I_{n_j}$, when the modes are completely persistent, and $P_{j\cdot} = \bar{p}'$ $(j \in N_j)$, when the modes are serially i.i.d. with probability distribution $\bar{p}$.

[8] Because mode-dependent intercepts (as well as mode-dependent standard deviations) are allowed in the model, we can still incorporate additive mode-dependent shocks.

with the modes, are the driving forces in the model and they are not directly observed. For technical reasons, it is convenient but not necessary that they are independent. We let $p_t = (p_{1t}, ..., p_{n_j t})'$ denote the true probability distribution of j_t in period t. We let $p_{t|t}$ denote the policymaker's and private sector's estimate of the probability distribution in the beginning of period t. The prediction equation for the probability distribution is

$$p_{t+1|t} = P' p_{t|t}. \tag{2.3}$$

We let the operator $\mathrm{E}_t[\cdot]$ in the expression $\mathrm{E}_t H_{j_{t+1}} x_{t+1}$ on the left side of (2.2) denote expectations in period t conditional on policymaker and private-sector information in the beginning of period t, including X_t, i_t, and $p_{t|t}$, but excluding j_t and ε_t. Thus, the maintained assumption is symmetric information between the policymaker and the (aggregate) private sector. Since forward-looking variables will be allowed to depend on j_t, *parts* of the private sector, but not the *aggregate* private sector, may be able to observe j_t and parts of ε_t. Note that although we focus on the determination of the optimal policy instrument i_t, our results also show how private sector choices as embodied in x_t are affected by uncertainty and learning. The precise informational assumptions and the determination of $p_{t|t}$ will be specified below.

We let the policymaker's intertemporal loss function in period t be

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

where δ is a discount factor satisfying $0 < \delta < 1$, and the period loss, $L(X_t, x_t, i_t, j_t)$, 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 ($j \in N_j$) is positive semidefinite. We assume that the policymaker optimizes under commitment in a timeless perspective. As explained below, we will then add the term

$$\Xi_{t-1} \frac{1}{\delta} \mathrm{E}_t H_{j_t} x_t \tag{2.6}$$

to the intertemporal loss function in period t. As we shall see below, the n_x-vector Ξ_{t-1} is the vector of Lagrange multipliers for equation (2.2) from the optimization problem in period $t-1$. For the special case when there are no forward-looking variables ($n_x = 0$), the model consists of (2.1) only, without the term $A_{12 j_{t+1}} x_t$; the period loss function depends on X_t, i_t, and j_t only; and there is no role for the Lagrange multipliers Ξ_{t-1} or the term (2.6).

We will distinguish three cases: (1) Optimal policy when there is no learning (NL), (2) Adaptive optimal policy (AOP), and (3) Bayesian optimal policy (BOP). By NL, we refer to a situation when the policymaker and the aggregate private sector have a probability distribution $p_{t|t}$ over the modes in period t and updates the probability distribution in future periods using the transition matrix only, so the updating equation is

$$p_{t+1|t+1} = P' p_{t|t}. \tag{2.7}$$

That is, the policymaker and the private sector do not use observations of the variables in the economy to update the probability distribution. The policymaker then determines optimal policy in period t conditional on $p_{t|t}$ and (2.7). This is a variant of a case examined in Svensson and Williams [15].

By AOP, we refer to a situation when the policymaker in period t determines optimal policy as in the NL case, but then uses observations of the realization of the variables in the economy to update its probability distribution according to Bayes Theorem. In this case, the instruments will generally have an effect on the updating of future probability distributions, and through this channel separately affect the intertemporal loss. However, the policymaker does not exploit that channel in determining optimal policy. That is, the policymaker does not do any conscious experimentation. By BOP, we refer to a situation when the policymaker acknowledges that the current instruments will affect future inference and updating of the probability distribution, and calculates optimal policy taking this separate channel into account. Therefore, BOP includes optimal experimentation, where for instance the policymaker may pursue policy that increases losses in the short run but improves the inference of the probability distribution and therefore lowers losses in the longer run.

3 Optimal policy with no learning

We first consider the NL case. Svensson and Williams [15] derive the equilibrium under commitment in a timeless perspective for the case when X_t, x_t, and i_t are observable in period t, j_t is unobservable, and the updating equation for $p_{t|t}$ is given by (2.7). Observations of X_t, x_t, and i_t are then not used to update $p_{t|t}$.

It is worth noting what type of belief specification underlies the assumption that the policymaker does not learn from his or her beliefs. In general this requires the policymaker to have subjective beliefs which are inconsistent or differ from the true data-generating process. One possibility would be to assume that the policymaker subjectively (and incorrectly) views modes as being independently drawn each period, in which case there is no reason to learn. As discussed in detail in Svensson and Williams [15], following a suggestion from Alexei Onatski, we instead assume that the policymaker in period t forgets past observations of the economy, such as $X_{t-1}, X_{t-2}, \ldots$, 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 can 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. This complication leads to the slightly more complicated derivations below.

As a further difference, Svensson and Williams [15] assumed $C_{2j_t} \equiv 0$. With observable modes or with unobservable modes with no learning, this is an innocuous assumption, since if $C_{2j_t} \not\equiv 0$ the vector of predetermined variables and the block of equations for the predetermined variables, (2.1), can be augmented with the vector $X_{\varepsilon t}$ and the equations $X_{\varepsilon,t+1} = C_{2j_{t+1}}\varepsilon_{t+1}$, respectively. Here we allow $C_{2j_t} \not\equiv 0$ and keep track of the term $C_{2j_t}\varepsilon_t$, since this term will serve as the shock in the equations for the forward-looking variables, without which inference in some cases becomes trivial.[9]

3.1 The general case

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

$$\mathrm{E}_t H_{j_{t+1}} x_{t+1} = z_t, \tag{3.1}$$

$$0 = A_{21j_t} X_t + A_{22j_t} x_t - z_t + B_{2j_t} i_t + C_{2j_t} \varepsilon_t, \tag{3.2}$$

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 (3.2) and solve x_t as a function of X_t, z_t, i_t, j_t, and ε_t

$$x_t = \tilde{x}(X_t, z_t, i_t, j_t, \varepsilon_t) \equiv A_{22j_t}^{-1}(z_t - A_{21j_t} X_t - B_{2j_t} i_t - C_{2j_t} \varepsilon_t). \tag{3.3}$$

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

In order to solve for the optimal decisions, we use the recursive saddlepoint method (see Marcet and Marimon [11], Svensson and Williams [15], and Svensson [14] for details of the recursive saddlepoint method). Thus, we introduce Lagrange multipliers for each forward looking equation, the

[9] Alternatively, we could allow $C_{2j_t} \equiv 0$ and add the corresponding predetermined variables, but then we have to assume that those predetermined variables are not observable. It turns out that the filtering problem becomes much more difficult to handle when some predetermined variables as well as modes are unobservable.

lagged values of which become state variables and reflecting costs of commitment, while the current values become control variables. The dual period loss function can be written

$$\mathrm{E}_t\tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j_t, \varepsilon_t) \equiv \sum_j p_{jt|t} \int \tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j, \varepsilon_t)\varphi(\varepsilon_t)d\varepsilon_t,$$

where $\tilde{X}_t \equiv (X_t', \Xi_{t-1}')'$ is the $(n_X + n_x)$-vector 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 $\varphi(\cdot)$ denotes a generic probability density function (for ε_t, the standard normal density function), and where

$$\tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j_t, \varepsilon_t) \equiv L[X_t, \tilde{x}(X_t, z_t, i_t, j_t, \varepsilon_t), i_t, j_t] - \gamma_t' z_t + \Xi_{t-1}'\frac{1}{\delta}H_{j_t}\tilde{x}(X_t, z_t, i_t, j_t, \varepsilon_t). \qquad (3.4)$$

As discussed in Svensson and Williams [15], the failure of the law of iterated expectations leads us to introduce the collection of value functions $\hat{V}(s_t, j)$ which condition on the mode, while the value function $\tilde{V}(s_t)$ averages over these and represents the solution of the dual optimization problem. The somewhat unusual Bellman equation for the dual problem can be written

$$\begin{aligned}
\tilde{V}(s_t) &= \mathrm{E}_t\hat{V}(s_t, j_t) \equiv \sum\nolimits_j p_{jt|t}\hat{V}(s_t, j) \\
&= \max_{\gamma_t} \min_{(z_t,i_t)} \mathrm{E}_t\{\tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j_t, \varepsilon_t) + \delta\hat{V}[g(s_t, z_t, i_t, \gamma_t, j_t, \varepsilon_t, j_{t+1}, \varepsilon_{t+1}), j_{t+1}]\} \\
&\equiv \max_{\gamma_t} \min_{(z_t,i_t)} \sum\nolimits_j p_{jt|t} \int \left[\begin{array}{l} \tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j, \varepsilon_t) \\ + \delta \sum_k P_{jk}\hat{V}[g(s_t, z_t, i_t, \gamma_t, j, \varepsilon_t, k, \varepsilon_{t+1}), k] \end{array} \right] \varphi(\varepsilon_t)\varphi(\varepsilon_{t+1})d\varepsilon_t d\varepsilon_{t+1}.
\end{aligned} \qquad (3.5)$$

where $s_t \equiv (\tilde{X}_t', p_{t|t}')'$ denotes the *perceived* state of the economy (it includes the perceived probability distribution, $p_{t|t}$, but not the true mode) and (s_t, j_t) denotes the *true* state of the economy (it includes the true mode of the economy). As we discuss in more detail below, 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 BOP case beliefs do satisfy this property, so the state vector is simply s_t. Also note that in the Bellman equation we require that all the choice variables respect the information constraints, and thus depend on the perceived state s_t but not the mode j directly.

The optimization is subject to the transition equation for X_t,

$$X_{t+1} = A_{11j_{t+1}}X_t + A_{12j_{t+1}}\tilde{x}(X_t, z_t, i_t, j_t, \varepsilon_t) + B_{1j_{t+1}}i_t + C_{1j_{t+1}}\varepsilon_{t+1}, \qquad (3.6)$$

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

$$\Xi_t = \gamma_t, \qquad (3.7)$$

and the transition equation (2.7) for $p_{t|t}$. Combining equations, we have the transition for s_t,

$$
\begin{aligned}
s_{t+1} \equiv \begin{bmatrix} X_{t+1} \\ \Xi_t \\ p_{t+1|t+1} \end{bmatrix} &= g(s_t, z_t, i_t, \gamma_t, j_t, \varepsilon_t, j_{t+1}, \varepsilon_{t+1}) \\
&\equiv \begin{bmatrix} A_{11j_{t+1}} X_t + A_{12j_{t+1}} \tilde{x}(X_t, z_t, i_t, j, \varepsilon_t) + B_{1j_{t+1}} i_t + C_{1j_{t+1}} \varepsilon_{t+1} \\ \gamma_t \\ P' p_{t|t} \end{bmatrix}. \qquad (3.8)
\end{aligned}
$$

It is straightforward to see that the solution of the dual optimization problem (3.5) is linear in $\tilde{X}_t$ for given $p_{t|t}, j_t$,

$$
\begin{bmatrix} z_t \\ i_t \\ \gamma_t \end{bmatrix} = \begin{bmatrix} z(s_t) \\ i(s_t) \\ \gamma(s_t) \end{bmatrix} = F(p_{t|t}) \tilde{X}_t \equiv \begin{bmatrix} F_z(p_{t|t}) \\ F_i(p_{t|t}) \\ F_\gamma(p_{t|t}) \end{bmatrix} \tilde{X}_t, \qquad (3.9)
$$

$$
x_t = x(s_t, j_t, \varepsilon_t) \equiv \tilde{x}(X_t, z(s_t), i(s_t), j_t, \varepsilon_t) \equiv F_{x\tilde{X}}(p_{t|t}, j_t) \tilde{X}_t + F_{x\varepsilon}(p_{t|t}, j_t) \varepsilon_t. \qquad (3.10)
$$

This solution is also the solution to the original primal optimization problem. We note that x_t is linear in ε_t for given $p_{t|t}$ and j_t. The equilibrium transition equation is then given by

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

As can be easily verified, the (unconditional) dual value function $\tilde{V}(s_t)$ is quadratic in $\tilde{X}_t$ for given $p_{t|t}$, taking the form

$$
\tilde{V}(s_t) \equiv \tilde{X}_t' \tilde{V}_{\tilde{X}\tilde{X}}(p_{t|t}) \tilde{X}_t + w(p_{t|t}).
$$

The conditional dual value function $\hat{V}(s_t, j_t)$ gives the dual intertemporal loss conditional on the true state of the economy, (s_t, j_t). It follows that this function satisfies

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

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

$$
\hat{V}(s_t, j_t) \equiv \tilde{X}_t' \hat{V}_{\tilde{X}\tilde{X}}(p_{t|t}, j_t) \tilde{X}_t + \hat{w}(p_{t|t}, j_t).
$$

It follows that we have

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

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 $\mathrm{E}_t \tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j_t, \varepsilon_t)$, satisfies

$$
\begin{aligned}
V(s_t) &\equiv \tilde{V}(s_t) - \Xi_{t-1}' \frac{1}{\delta} \sum_j p_{jt|t} H_j \int x(s_t, j, \varepsilon_t) \varphi(\varepsilon_t) d\varepsilon_t \\
&= \tilde{V}(s_t) - \Xi_{t-1}' \frac{1}{\delta} \sum_j p_{jt|t} H_j x(s_t, j, 0) \qquad (3.11)
\end{aligned}
$$

(where the second equality follows since $x(s_t, j_t, \varepsilon_t)$ is linear in ε_t for given s_t and j_t). It is quadratic in $\tilde{X}_t$ for given $p_{t|t}$,

$$V(s_t) \equiv \tilde{X}_t' V_{\tilde{X}\tilde{X}}(p_{t|t})\tilde{X}_t + w(p_{t|t})$$

(the scalar $w(p_{t|t})$ 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 $p_{t|t}$ after X_t has been observed but before x_t has been observed, taking into account that j_t and ε_t are not observed. Hence, the second term on the right side of (3.11) contains the expectation of $H_{j_t} x_t$ conditional on that information.[10]

Svensson and Williams [15] present algorithms to compute the solution and the primal and dual value functions for the no-learning case, with and without forward-looking variables, when the matrices $C_{2j} \equiv 0$. For completeness, appendix A presents variants of these algorithms that incorporate the case when $C_{2j_t} \not\equiv 0$. For future reference, we note that the value function for the primal problem also satisfies

$$V(s_t) \equiv \sum\nolimits_j p_{jt|t} \check{V}(s_t, j),$$

where the conditional value function, $\check{V}(s_t, j_t)$, satisfies

$$\check{V}(s_t, j) = \int \left\{ \begin{array}{l} L[X_t, x(s_t, j, \varepsilon_t), i(s_t), j] \\ +\,\delta \sum_k P_{jk} \check{V}[\bar{g}(s_t, j, \varepsilon_t, k, \varepsilon_{t+1}), k] \end{array} \right\} \varphi(\varepsilon_t)\varphi(\varepsilon_{t+1}) d\varepsilon_t d\varepsilon_{t+1} \qquad (j \in N_j). \tag{3.12}$$

3.2 The case without forward-looking variables

For the case without forward-looking variables, the recursive saddlepoint method is not needed, so matters simplify. The transition equation for X_{t+1} is

$$X_{t+1} = A_{j_{t+1}} X_t + B_{j_{t+1}} i_t + C_{j_{t+1}} \varepsilon_{t+1}, \tag{3.13}$$

and the period loss function is

$$\mathrm{E}_t L(X_t, i_t, j_t) \equiv \sum\nolimits_j p_{jt|t} L(X_t, i_t, j), \tag{3.14}$$

where

$$L(X_t, i_t, j_t) \equiv \begin{bmatrix} X_t \\ i_t \end{bmatrix}' W_{j_t} \begin{bmatrix} X_t \\ i_t \end{bmatrix}. \tag{3.15}$$

The transition equation is

$$s_{t+1} \equiv \begin{bmatrix} X_{t+1} \\ p_{t+1|t+1} \end{bmatrix} = g(s_t, i_t, j_{t+1}, \varepsilon_{t+1}) \equiv \begin{bmatrix} A_{j_{t+1}} X_t + B_{j_{t+1}} i_t + C_{j_{t+1}} \varepsilon_{t+1} \\ P' p_{t|t} \end{bmatrix}. \tag{3.16}$$

[10] To be precise, the observation of X_t, which depends on $C_{1j_t}\varepsilon_t$, allows some inference of ε_t, $\varepsilon_{t|t}$. x_t will depend on j_t and on ε_t, but on ε_t only through $C_{2j_t}\varepsilon_t$. By assumption $C_{1j}\varepsilon_t$ and $C_{2k}\varepsilon_t$ are independent. Hence, any observation of X_t and $C_{1j}\varepsilon_t$ does not convey any information about $C_{2j}\varepsilon_t$, so $\mathrm{E}_t C_{2j_t}\varepsilon_t = 0$.

The Bellman equation for the optimal policy problem is

$$
\begin{aligned}
V(s_t) &= \mathrm{E}_t \hat{V}(s_t, j_t) \equiv \sum\nolimits_j p_{jt|t} \hat{V}(s_t, j) \\
&= \min_{i_t} \mathrm{E}_t \{ L(X_t, i_t, j_t) + \delta \hat{V}[g(s_t, i_t, j_{t+1}, \varepsilon_{t+1}), j_{t+1}] \} \\
&\equiv \min_{i_t} \sum\nolimits_j p_{jt|t} \left[L(X_t, i_t, j) + \delta \sum\nolimits_k P_{jk} \int \hat{V} \left[g(s_t, i_t, k, \varepsilon_{t+1}), k \right] \varphi(\varepsilon_{t+1}) d\varepsilon_{t+1} \right]. \qquad (3.17)
\end{aligned}
$$

This results in the optimal policy function,

$$
i_t = i(s_t) \equiv F_i(p_{t|t}) X_t, \qquad (3.18)
$$

which is linear in X_t for given $p_{t|t}$. The equilibrium transition equation is then

$$
s_{t+1} = \bar{g}(s_t, j_{t+1}, \varepsilon_{t+1}) \equiv g(s_t, i(s_t), j_{t+1}, \varepsilon_{t+1}). \qquad (3.19)
$$

The value function, $V(s_t)$, is quadratic in X_t for given $p_{t|t}$,

$$
V(s_t) = X_t' V_{XX}(p_{t|t}) X_t + w(p_{t|t}).
$$

The conditional value function, $\hat{V}(s_t, j_t)$, satisfies

$$
\hat{V}(s_t, j) \equiv L[X_t, i(s_t), j] + \delta \sum\nolimits_k P_{jk} \int \hat{V}[\bar{g}(s_t, k, \varepsilon_{t+1}), k] \varphi(\varepsilon_{t+1}) d\varepsilon_{t+1} \qquad (j \in N_j).
$$

4 Adaptive optimal policy

Consider now the case of adaptive optimal policy, where the policymaker uses the same policy function as in the no-learning case, but each period updates the probabilities that this policy is conditioned on. This case is thus simple to implement recursively, as we have already discussed how to solve for the optimal decisions and below we show how to update probabilities. However, the ex-ante evaluation of expected loss is more complex, as we show below. In particular, we assume that $C_{2j_t} \not\equiv 0$ and that both ε_t and j_t are unobservable. The estimate $p_{t|t}$ is the result of Bayesian updating, using all information available, but the optimal policy in period t is computed under the perceived updating equation (2.7). That is, the fact that the policy choice will affect future $p_{t+\tau|t+\tau}$ *and* that future expected loss will change when $p_{t+\tau|t+\tau}$ changes is disregarded. Under the assumption that the expectations on the left side of (2.2) are conditional on (2.7), the variables z_t, i_t, γ_t, and x_t in period t are still determined by (3.9) and (3.10).

4.1 Information revelation

In order to determine the updating equation for $p_{t|t}$, we specify an explicit sequence of information revelation as follows, in no less than nine steps. The timing assumptions are necessary in order to spell out the appropriate conditioning for decisions and updating of beliefs.

First, the policymaker and the private sector enters period t with the prior $p_{t|t-1}$. They know X_{t-1}, $x_{t-1} = x(s_{t-1}, j_{t-1}, \varepsilon_{t-1})$, $z_{t-1} = z(s_{t-1})$, $i_{t-1} = i(s_{t-1})$, and $\Xi_{t-1} = \gamma(s_{t-1})$ from the previous period.

Second, in the beginning of period t, the mode j_t and the vector of shocks ε_t are realized. Then the vector of predetermined variables X_t is realized according to (2.1).

Third, the policymaker and the private sector observe X_t. They then know $\tilde{X}_t \equiv (X_t', \Xi_{t-1}')'$. They do not observe j_t or ε_t

Fourth, the policymaker and the private sector update the prior $p_{t|t-1}$ to the posterior $p_{t|t}$ according to Bayes Theorem and the updating equation

$$p_{jt|t} = \frac{\varphi(X_t | j_t = j, X_{t-1}, x_{t-1}, i_{t-1}, p_{t|t-1})}{\varphi(X_t | X_{t-1}, x_{t-1}, i_{t-1}, p_{t|t-1})} p_{jt|t-1} \qquad (j \in N_j), \tag{4.1}$$

where again $\varphi(\cdot)$ denotes a generic density function.[11] Then the policymaker and the private sector know $s_t \equiv (\tilde{X}_t', p_{t|t}')'$.

Fifth, the policymaker solves the dual optimization problem, determines $i_t = i(s_t)$, and implements/announces the instrument setting i_t.

Sixth, the private-sector (and policymaker) expectations,

$$z_t = \mathrm{E}_t H_{j_{t+1}} x_{t+1} \equiv \mathrm{E}[H_{j_{t+1}} x_{t+1} \,|\, s_t],$$

are formed. In equilibrium, these expectations will be determined by (3.9). In order to understand their determination better, we look at this in some detail.

These expectations are by assumption formed before x_t is observed. The private sector and the policymaker know that x_t will in equilibrium be determined in the next step according to (3.10). Hence, they can form expectations of the soon-to-be determined x_t conditional on $j_t = j$,[12]

$$x_{jt|t} = x(s_t, j, 0). \tag{4.2}$$

[11] The policymaker and private sector can also estimate the shocks $\varepsilon_{t|t}$ as $\varepsilon_{t|t} = \sum_j p_{jt|t} \varepsilon_{jt|t}$, where $\varepsilon_{jt|t} \equiv X_t - A_{11j} X_{t-1} - A_{12j} x_{t-1} - B_{1j} i_{t-1}$ $(j \in N_j)$. However, because of the assumed independence of $C_{1j}\varepsilon_t$ and $C_{2k}\varepsilon_t$, $j, k \in N_j$, we do not need to keep track of $\varepsilon_{jt|t}$.

[12] Note that 0 instead of $\varepsilon_{jt|t}$ enters above. This is because the inference $\varepsilon_{jt|t}$ above is inference about $C_{1j}\varepsilon_t$, whereas x_t depends on ε_t through $C_{2j}\varepsilon_t$. Since we assume that $C_{1j}\varepsilon_t$ and $C_{2j}\varepsilon_t$ are independent, there is no inference of $C_{2j}\varepsilon_t$ from observing X_t. Hence, $\mathrm{E}_t C_{2j_t} \varepsilon_t \equiv 0$. Because of the linearity of x_t in ε_t, the integration of x_t over ε_t results in $x(s_t, j_t, 0_t)$.

The private sector and the policymaker can also infer Ξ_t from

$$\Xi_t = \gamma(s_t). \tag{4.3}$$

This allows the private sector and the policymaker to form the expectations

$$z_t = z(s_t) = \mathrm{E}_t[H_{j_{t+1}} x_{t+1} \mid s_t] = \sum\nolimits_{j,k} P_{jk} p_{jt|t} H_k x_{k,t+1|jt}, \tag{4.4}$$

where

$$\begin{aligned} x_{k,t+1|jt} &= \int x\left(\begin{bmatrix} A_{11k}X_t + A_{12k}x(s_t, j, \varepsilon_t) + B_{1k}i(s_t) \\ \Xi_t \\ P' p_{t|t} \end{bmatrix}, k, \varepsilon_{t+1}\right) \varphi(\varepsilon_t)\varphi(\varepsilon_{t+1}) d\varepsilon_t d\varepsilon_{t+1} \\ &= x\left(\begin{bmatrix} A_{11k}X_t + A_{12k}x(s_t, j, 0) + B_{1k}i(s_t) \\ \Xi_t \\ P' p_{t|t} \end{bmatrix}, k, 0\right), \end{aligned}$$

where we have exploited the linearity of $x_t = x(s_t, j_t, \varepsilon_t)$ and $x_{t+1} = x(s_{t+1}, j_{t+1}, \varepsilon_{t+1})$ in ε_t and ε_{t+1}. Note that z_t is, under AOP, formed conditional on the belief that the probability distribution in period $t+1$ will be given by $p_{t+1|t+1} = P' p_{t|t}$, not by the true updating equation that we are about to specify.

Seventh, after the expectations z_t have been formed, x_t is determined as a function of X_t, z_t, i_t, j_t, and ε_t by (3.3).

Eight, the policymaker and the private sector then use the observed x_t to update $p_{t|t}$ to the new posterior $p^+_{t|t}$ according to Bayes Theorem, via the updating equation

$$p^+_{jt|t} = \frac{\varphi(x_t | j_t = j, X_t, z_t, i_t, p_{t|t})}{\varphi(x_t | X_t, z_t, i_t, p_{t|t})} p_{jt|t} \qquad (j \in N_j). \tag{4.5}$$

Ninth, the policymaker and the private sector then leave period t and enter period $t+1$ with the prior $p_{t+1|t}$ given by the prediction equation

$$p_{t+1|t} = P' p^+_{t|t}. \tag{4.6}$$

In the beginning of period $t+1$, the mode j_{t+1} and the vector of shocks ε_{t+1} are realized, and X_{t+1} is determined by (2.1) and observed by the policymaker and private sector. The sequence of the nine steps above then repeats itself.

With the timing laid out, we now provide more detail on the updating equations (4.1) and (4.5), explicitly writing out the densities. This will help in writing an explicit law of motion for beliefs.

Since $C_{1j}\varepsilon_t$ is a random n_X-vector that, for given j, is normally distributed with mean zero and covariance matrix $C_{1j}C_{1j}'$,[13] we know that

$$\varphi(X_t|j_t = j, X_{t-1}, x_{t-1}, i_{t-1}, p_{t|t-1}) \equiv \psi(X_t - A_{11j}X_{t-1} - A_{12j}x_{t-1} - B_{1j}i_{t-1}; C_{1j}C_{1j}'), \quad (4.7)$$

where

$$\psi(\varepsilon; \Sigma_{\varepsilon\varepsilon}) \equiv \frac{1}{\sqrt{(2\pi)^{n_\varepsilon} |\Sigma_{\varepsilon\varepsilon}|}} \exp\left(-\frac{1}{2}\varepsilon'\Sigma_{\varepsilon\varepsilon}^{-1}\varepsilon\right)$$

denotes the density function of a random n_ε-vector ε with a multivariate normal distribution with mean zero and covariance matrix $\Sigma_{\varepsilon\varepsilon}$. Furthermore,

$$\varphi(X_t|X_{t-1}, x_{t-1}, i_{t-1}, p_{t|t-1}) \equiv \sum_j p_{jt|t-1}\psi(X_t - A_{11j}X_{t-1} + A_{12j}x_{t-1} + B_{1j}i_{t-1}; C_{1j}C_{1j}'). \quad (4.8)$$

Thus, we know the details of the updating equation (4.1). Further, since $C_{2k}\varepsilon_t$ is a random n_x-vector that is normally distributed with mean zero and covariance matrix $C_{2k}C_{2k}'$, we know that[14]

$$\varphi(x_t|j_t = k, X_t, z_t, i_t, p_{t|t}) \equiv \psi[z_t - A_{21k}X_t - A_{22k}x_t - B_{2k}i_t; C_{2k}C_{2k}'], \quad (4.9)$$

$$\varphi(x_t|X_t, z_t, i_t, p_{t|t}) \equiv \sum_k p_{kt|t}\psi[z_t - A_{21k}X_t - A_{22k}x_t - B_{2k}i_t; C_{2k}C_{2k}']. \quad (4.10)$$

Thus, we know the details of the updating equation (4.5).

In particular, it follows that we can write the updating equation (4.5) as

$$\begin{aligned} p_{t|t}^+ &= Q^+(s_t, z_t, i_t, j_t, \varepsilon_t) \quad (4.11) \\ &\equiv [Q_1^+(s_t, z_t, i_t, j_t, \varepsilon_t), ..., Q_{n_j}^+(s_t, z_t, i_t, j_t, \varepsilon_t)]', \end{aligned}$$

where

$$Q_k^+(s_t, z_t, i_t, j_t, \varepsilon_t) \equiv \frac{\psi[Z_k(X_t, z_t, i_t, j_t, \varepsilon_t); C_{2k}C_{2k}']}{\sum_k p_{kt|t}\psi[Z_k(X_t, z_t, i_t, j_t, \varepsilon_t); C_{2k}C_{2k}']} p_{kt|t} \qquad (k \in N_j) \quad (4.12)$$

and

$$Z_k(X_t, z_t, i_t, j_t, \varepsilon_t) \equiv z_t - A_{21k}X_t - A_{22k}\tilde{x}(X_t, z_t, i_t, j_t, \varepsilon_t) - B_{2k}i_t,$$

where we use (3.3) to express x_t as a function of X_t, z_t, i_t, j_t, and ε_t, and use this to eliminate x_t from the first argument of $\psi(\cdot, \cdot)$ in (4.9) and (4.10).

The transition equation for $p_{t+1|t+1}$ can then finally be written

$$p_{t+1|t+1} = Q(s_t, z_t, i_t, j_t, \varepsilon_t, j_{t+1}, \varepsilon_{t+1}), \quad (4.13)$$

[13] Assume for simplicity that the rank of $C_{1j}C_{1j}'$ is n_X; if not, for instance when the predetermined variables include lagged endogenous variables, choose the appropriate nonsingular submatrix and the appropriate subvector of X_t.

[14] Again, assume that the rank of $C_{2k}C_{2k}'$ is n_x, or else select the appropriate nonsingular components.

where $Q(s_t, z_t, i_t, j_t, \varepsilon_t, j_{t+1}, \varepsilon_{t+1})$ is defined by the combination of (4.1) for period $t+1$ with (3.6), (4.6), and (4.11). The equilibrium transition equation for the full state vector is then given by

$$s_{t+1} \equiv \begin{bmatrix} X_{t+1} \\ \Xi_t \\ p_{t+1|t+1} \end{bmatrix} = \bar{g}(s_t, j_t, \varepsilon_t, j_{t+1}, \varepsilon_{t+1})$$
$$\equiv \begin{bmatrix} A_{11j_{t+1}} X_t + A_{12j_{t+1}} x(s_t, j_t, \varepsilon_t) + B_{1j_{t+1}} i(s_t) + C_{1j_{t+1}} \varepsilon_{t+1} \\ \gamma(s_t) \\ Q(s_t, z(s_t), i(s_t), j_t, \varepsilon_t, j_{t+1}, \varepsilon_{t+1}) \end{bmatrix}, \qquad (4.14)$$

where the third row is given by the true updating equation (4.13) together with the policy function (3.9). Thus, we note that in this AOP case there is a distinction between the "perceived" transition equation, which includes the perceived updating equation, (2.7), and the "true" transition equation, which includes the true updating equation (4.13).

Note that $V(s_t)$ in (3.11), which is subject to the perceived transition equation, (3.8), does not give the true (unconditional) value function for the AOP case. This is instead given by

$$\bar{V}(s_t) \equiv \sum\nolimits_j p_{jt|t} \check{V}(s_t, j),$$

where the true conditional value function, $\check{V}(s_t, j_t)$, satisfies

$$\check{V}(s_t, j) = \int \left\{ \begin{array}{l} L[X_t, x(s_t, j, \varepsilon_t), i(s_t), j] \\ + \delta \sum_k P_{jk} \check{V}[\bar{g}(s_t, j, \varepsilon_t, k, \varepsilon_{t+1}), k] \end{array} \right\} \varphi(\varepsilon_t) \varphi(\varepsilon_{t+1}) d\varepsilon_t d\varepsilon_{t+1} \qquad (j \in N_j). \qquad (4.15)$$

That is, the true value function $\bar{V}(s_t)$ takes into account the true updating equation for $p_{t|t}$, (4.13), whereas the optimal policy, (3.9), and the perceived value function, $V(s_t)$ in (3.11), are conditional on the perceived updating equation (2.7) and thereby the perceived transition equation (3.8). Note also that $\bar{V}(s_t)$ is the value function after $\tilde{X}_t$ has been observed but before x_t is observed, so it is conditional on $p_{t|t}$ rather than $p^+_{t|t}$. Since the full transition equation (4.14) is no longer linear due to the belief updating (4.13), the true value function $\bar{V}(s_t)$ is no longer quadratic in $\tilde{X}_t$ for given $p_{t|t}$. Thus, more complex numerical methods are required to evaluate losses in the AOP case, although policy is still determined simply as in the NL case.

Note that

$$\mathrm{E}_t Q(s_t, z(s_t), i(s_t), j_t, \varepsilon_t, j_{t+1}, \varepsilon_{t+1}) = p_{t+1|t} = P' p_{t|t}. \qquad (4.16)$$

The difference between the true updating equation for $p_{t+1|t+1}$, (4.13), and the perceived updating equation (2.7) is that, in the true updating equation, $p_{t+1|t+1}$ becomes a random variable from the point of view of period t, with mean equal to $p_{t+1|t}$. This is because $p_{t+1|t+1}$ depends on the realization of j_{t+1} and ε_{t+1}. We can hence write the true transition equation for $p_{t+1|t+1}$ as

$$p_{t+1|t+1} = P' p_{t|t} + v_{t+1} \qquad (4.17)$$

where

$$v_{t+1} \equiv Q(s_t, z(s_t), i(s_t), j_t, \varepsilon_t, j_{t+1}, \varepsilon_{t+1}) - P' p_{t|t}, \tag{4.18}$$

and thus $E_t v_{t+1} = 0$. The first term on the right side of (4.17) is the prediction $p_{t+1|t}$ and the second term is the innovation in $p_{t+1|t+1}$ that results from the Bayesian updating and depends on the realization of j_{t+1} and ε_{t+1}.

This way of representing beliefs sheds light on the gains from learning. If the conditional value function $\check{V}(s_t, j_t)$ under NL is concave in $p_{t|t}$ for given $\tilde{X}_t$ and j_t, then by Jensen's inequality the true expected future loss under AOP will be lower than the true expected future loss under NL. That is, the concavity of the value function in beliefs means that learning leads to lower losses. While it likely that $\check{V}$ is indeed concave, as we show in applications, it need not be globally so and thus learning need not always reduce losses. In some cases the losses incurred by increased variability of beliefs may offset the expected precision gains. Furthermore, under BOP, it may be possible to adjust policy so as to further increase the variance of $p_{t|t}$, that is, achieve a mean-preserving spread which might further reduce the expected future loss.[15] This amounts to optimal experimentation.

4.2 The case without forward-looking variables

For the case without forward-looking variables, again the recursive saddlepoint method is not needed and the expressions simplify. With the transition equation for the predetermined variables (3.13) and the period loss function (3.14), the optimal policy in the AOP case is determined as in the NL case by the solution to (3.17), subject to the perceived transition equation (3.16) and given by the same policy function, (3.18).

The optimal policy under AOP is calculated under the perceived updating equation, (2.7). The true updating equation for $p_{t+1|t+1}$ is

$$p_{t+1|t+1} = Q(s_t, i_t, j_{t+1}, \varepsilon_{t+1}), \tag{4.19}$$

where

$$Q(s_t, i_t, j_{t+1}, \varepsilon_{t+1}) \equiv [Q_1(s_t, i_t, j_{t+1}, \varepsilon_{t+1}), ..., Q_{n_j}(s_t, i_t, j_{t+1}, \varepsilon_{t+1})]',$$

$$Q_k(s_t, i_t, j_{t+1}, \varepsilon_{t+1}) \equiv$$
$$\frac{\psi[(A_{j_{t+1}} - A_k)X_t + (B_{j_{t+1}} - B_k)i_t + C_{j_{t+1}}\varepsilon_{t+1}; C_k C_k']}{\sum_{j,k} P_{jk} p_{jt|t} \psi[(A_{j_{t+1}} - A_k)X_t + (B_{j_{t+1}} - B_k)i_t + C_{j_{t+1}}\varepsilon_{t+1}; C_k C_k']} \sum_j P_{jk} p_{jt|t}.$$

[15] Kiefer [9] examines the properties of a value function, including concavity, under Bayesian learning for a simpler model without forward looking variables.

The equilibrium transition equation is then

$$s_{t+1} = \bar{g}(s_t, j_{t+1}, \varepsilon_{t+1}) \equiv \begin{bmatrix} A_{j_{t+1}} X_t + B_{j_{t+1}} i(s_t) + C_{j_{t+1}} \varepsilon_{t+1} \\ Q(s_t, i(s_t), j_{t+1}, \varepsilon_{t+1}) \end{bmatrix}.$$

The true (unconditional) value function, $\bar{V}(s_t)$, taking into account that $p_{t+1|t+1}$ will be updated according to (4.19) and ex post depend on j_{t+1} and ε_{t+1}, is given by

$$\bar{V}(s_t) \equiv \sum\nolimits_j p_{jt|t} \check{V}(s_t, j),$$

where the true conditional value function $\check{V}(s_t, j)$ satisfies

$$\check{V}(s_t, j) = L[X_t, i(s_t), j] + \delta \sum\nolimits_k P_{jk} \int \check{V}[\bar{g}(s_t, k, \varepsilon_{t+1}), k] \varphi(\varepsilon_{t+1}) d\varepsilon_{t+1}.$$

Again, if the conditional value function $\check{V}(s_{t+1}, j_{t+1})$ under NL is concave in $p_{t+1|t+1}$, the value function $\bar{V}(s_t)$ under AOP will be lower than under NL.

4.3 A special case when forward-looking variables do not reveal any further information

A special case that is simpler to deal with is when the forward-looking equation (2.2) does not vary with the mode:

$$A_{21j} = A_{21}, \quad A_{22j} = A_{22}, \quad B_{2j} = B_2, \quad C_{2j} = 0 \quad (j \in N_j). \tag{4.20}$$

That is, the matrices A_{21}, A_{22}, and B_2 are independent of j, and the matrix $C_2 = 0$, so

$$x_t = \tilde{x}(X_t, z_t, i_t) \equiv A_{22}^{-1}(z_t - A_{21} X_t - B_2 i_t).$$

In that case, the observation of x_t does not reveal any further information about j_t. This implies that the updating equation (4.5) collapses to

$$p_{t|t}^{+} = p_{t|t},$$

so the prediction equation (4.6) is simply

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

In particular, we then have

$$\begin{aligned} x_t = x(s_t) \;&\equiv\; \tilde{x}[X_t, z(s_t), i(s_t)], \\ p_{t+1|t+1} &= Q(s_t, z_t, i_t, j_{t+1}, \varepsilon_{t+1}), \\ s_{t+1} &= g(s_t, z_t, i_t, \gamma_t, j_{t+1}, \varepsilon_{t+1}), \\ \bar{g}(s_t, j_{t+1}, \varepsilon_{t+1}) &\equiv g(s_t, z(s_t), i(s_t), \gamma(s_t), j_{t+1}, \varepsilon_{t+1}). \end{aligned}$$

That is, there is in this case no separate dependence of s_{t+1} and x_t on j_t and ε_t beyond s_t. This special case also makes the Bayesian optimal policy simpler, as we see below.

5 Bayesian optimal policy

Finally, we consider the BOP case, when optimal policy is determined while taking the updating equation (4.13) into account. That is, we now allow the policymaker to choose i_t taking into account that his actions will affect $p_{t+1|t+1}$, which in turn will affect future expected losses. In particular, experimentation is allowed and is optimally chosen. For the BOP case, there is hence no distinction between the "perceived" and "true" transition equation.

5.1 The general case

The transition equation for the BOP case is:

$$\begin{aligned} s_{t+1} &\equiv \begin{bmatrix} X_{t+1} \\ \Xi_t \\ p_{t+1|t+1} \end{bmatrix} = g(s_t, z_t, i_t, \gamma_t, j_t, \varepsilon_t, j_{t+1}, \varepsilon_{t+1}) \\ &\equiv \begin{bmatrix} A_{11j_{t+1}} X_t + A_{12j_{t+1}} \tilde{x}(s_t, z_t, i_t, j_t, \varepsilon_t) + B_{1j_{t+1}} i_t + C_{1j_{t+1}} \varepsilon_{t+1} \\ \gamma_t \\ Q(s_t, z_t, i_t, j_t, \varepsilon_t, j_{t+1}, \varepsilon_{t+1}) \end{bmatrix}. \end{aligned} \tag{5.1}$$

Then the dual optimization problem can be written as (3.5) subject to the above transition equation (5.1). However, in the Bayesian case, matters simplify somewhat, as we do not need to compute the conditional value functions $\hat{V}(s_t, j_t)$, which we recall were required due to the failure of the law of iterated expectations in the AOP case. We note now that the second term on the right side of (3.5) can be written as

$$\mathrm{E}_t \hat{V}(s_{t+1}, j_{t+1}) \equiv \mathrm{E}\left[\hat{V}(s_{t+1}, j_{t+1}) \,\middle|\, s_t \right].$$

Since, in the Bayesian case, the beliefs do satisfy the law of iterated expectations, this is then the same as

$$
\begin{aligned}
\mathrm{E}\left[\hat{V}(s_{t+1}, j_{t+1}) \,\middle|\, s_t\right] &= \mathrm{E}\left[\hat{V}\left(\left[\begin{array}{c} X_{t+1}(j_{t+1}, \varepsilon_{t+1}) \\ \Xi_t \\ p_{t+1|t+1}(X_{t+1}(j_{t+1}, \varepsilon_{t+1})) \end{array}\right], j_{t+1}\right) \,\middle|\, s_t\right] \\
&= \mathrm{E}\left\{\mathrm{E}\left[\hat{V}\left(\left[\begin{array}{c} X_{t+1}(j_{t+1}, \varepsilon_{t+1}) \\ \Xi_t \\ p_{t+1|t+1}(X_{t+1}(j_{t+1}, \varepsilon_{t+1})) \end{array}\right], j_{t+1}\right) \,\middle|\, X_{t+1}, p_{t+1|t+1}(X_{t+1})\right] \,\middle|\, s_t\right\} \\
&= \mathrm{E}\left[\tilde{V}\left(\left[\begin{array}{c} X_{t+1}(j_{t+1}, \varepsilon_{t+1}) \\ \Xi_t \\ p_{t+1|t+1}(X_{t+1}(j_{t+1}, \varepsilon_{t+1})) \end{array}\right]\right) \,\middle|\, s_t\right] \\
&= \mathrm{E}\left[\tilde{V}(s_{t+1}) \,\middle|\, s_t\right],
\end{aligned}
$$

where we use the definition of $\tilde{V}(s_t)$, that X_{t+1} is a function of j_{t+1} and ε_{t+1}, and that $p_{t+1|t+1}$ is a function of X_{t+1}. Appendix B provides a more detailed proof.

Thus, the dual Bellman equation for the Bayesian optimal policy is

$$
\begin{aligned}
\tilde{V}(s_t) &= \max_{\gamma_t} \min_{(z_t, i_t)} \mathrm{E}_t\{\tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j_t, \varepsilon_t) + \delta\tilde{V}[g(s_t, z_t, i_t, \gamma_t, j_t, \varepsilon_t, j_{t+1}, \varepsilon_{t+1})]\} \\
&\equiv \max_{\gamma_t} \min_{(z_t, i_t)} \sum\nolimits_j p_{jt|t} \int \left[\begin{array}{l} \tilde{L}(\tilde{X}_t, z_t, i_t, \gamma_t, j, \varepsilon_t) \\ +\,\delta \sum_k P_{jk}\tilde{V}[g(s_t, z_t, i_t, \gamma_t, j, \varepsilon_t, k, \varepsilon_{t+1})] \end{array}\right] \varphi(\varepsilon_t)\varphi(\varepsilon_{t+1})d\varepsilon_t d\varepsilon_{t+1},
\end{aligned} \tag{5.2}
$$

where the transition equation is given by (5.1).

The solution to the optimization problem can be written

$$
\tilde{\imath}_t \equiv \left[\begin{array}{c} z_t \\ i_t \\ \gamma_t \end{array}\right] = \tilde{\imath}(s_t) \equiv \left[\begin{array}{c} z(s_t) \\ i(s_t) \\ \gamma(s_t) \end{array}\right] = F(\tilde{X}_t, p_{t|t}) \equiv \left[\begin{array}{c} F_z(\tilde{X}_t, p_{t|t}) \\ F_i(\tilde{X}_t, p_{t|t}) \\ F_\gamma(\tilde{X}_t, p_{t|t}) \end{array}\right], \tag{5.3}
$$

$$
x_t = x(s_t, j_t, \varepsilon_t) \equiv \tilde{x}(X_t, z(s_t), i(s_t), j_t, \varepsilon_t) \equiv F_x(\tilde{X}_t, p_{t|t}, j_t, \varepsilon_t). \tag{5.4}
$$

Because of the nonlinearity of (4.13) and (5.1), the solution is no longer linear in $\tilde{X}_t$ for given $p_{t|t}$. The dual value function, $\tilde{V}\,(s_t)$, is no longer quadratic in $\tilde{X}_t$ for given $p_{t|t}$. The value function of the primal problem, $V(s_t)$, is given by, equivalently, (3.11), (4.15) (with the equilibrium transition equation (4.14) with the solution (5.3)), or

$$
V(s_t) = \sum_j p_{jt|t} \int \left\{\begin{array}{l} L[X_t, x(s_t, j, \varepsilon_t), i(s_t), j] \\ +\,\delta \sum_k P_{jk} V[\bar{g}(s_t, j, \varepsilon_t, k, \varepsilon_{t+1})] \end{array}\right\} \varphi(\varepsilon_t)\varphi(\varepsilon_{t+1})d\varepsilon_t d\varepsilon_{t+1}. \tag{5.5}
$$

It it is also no longer quadratic in $\tilde{X}_t$ for given $p_{t|t}$. Thus, more complex and detailed numerical methods are necessary in this case to find the optimal policy and the value function. Therefore little can be said in general about the solution of the problem. Nonetheless, in numerical analysis it is very useful to have a good starting guess at a solution, which in our case comes from the AOP case. In our examples below we explain in more detail how the BOP and AOP cases differ, and what drives the differences.

5.2 The case without forward-looking variables

In the case without forward-looking variables, the transition equation for $s_{t+1|t+1}$ is

$$s_{t+1} = g(s_t, i_t, j_{t+1}, \varepsilon_{t+1}) \equiv \begin{bmatrix} A_{j_{t+1}} X_t + B_{j_{t+1}} i_t + C_{j_{t+1}} \varepsilon_{t+1} \\ Q(s_t, i_t, j_{t+1}, \varepsilon_{t+1}) \end{bmatrix},$$

and the optimal policy is determined by the Bellman equation

$$\begin{aligned} V(s_t) &= \min_{i_t} \mathrm{E}_t\{[L(X_t, i(s_t), j_t) + \delta V[g(s_t, i_t, j_{t+1}, \varepsilon_{t+1})]\} \\ &= \min_{i_t} \sum\nolimits_j p_{jt|t} \left\{ L(X_t, i_t, j) + \delta \sum\nolimits_k P_{jk} \int V[g(s_t, i_t, k, \varepsilon_{t+1})] \varphi(\varepsilon_{t+1}) d\varepsilon_{t+1} \right\}. \end{aligned}$$

This results in the optimal policy function

$$i_t = i(s_t) \equiv F_i(s_t).$$

Again, because of the nonlinearity of $Q(s_t, i_t, j_{t+1}, \varepsilon_{t+1})$, the optimal policy function is no longer linear in X_t for given $p_{t|t}$, and the value function is no longer quadratic in X_t for given $p_{t|t}$. The equilibrium transition equation is

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

5.3 The special case when forward-looking variables do not reveal any further information

As above, the special case (4.20) makes it unnecessary to deal with the details of the updating equation (4.11) and the separate dependence of s_{t+1} on j_t and ε_t. The transition equation is simply

$$\begin{aligned} s_{t+1} \equiv \begin{bmatrix} X_{t+1} \\ \Xi_t \\ p_{t+1|t+1} \end{bmatrix} &= g(s_t, z_t, i_t, \gamma_t, j_{t+1}, \varepsilon_{t+1}) \\ &\equiv \begin{bmatrix} A_{11j_{t+1}} X_t + A_{12j_{t+1}} \tilde{x}(s_t, z_t, i_t) + B_{1j_{t+1}} i_t + C_{1j_{t+1}} \varepsilon_{t+1} \\ \gamma_t \\ Q(s_t, z_t, i_t, j_{t+1}, \varepsilon_{t+1}) \end{bmatrix}. \end{aligned}$$

5.4 Bayesian optimal policy with endogenous mode transition

In the baseline formulation of the model, the mode transition matrix is given, so the model uncertainty represented by the Markov chain of the modes is independent of the state of the economy and the policy choice. However, in some situations it is natural to think of the state or the policy as influencing the likelihoods of the different modes. Assume now that the mode transition probabilities are instead endogenous and do depend on X_t, x_t, and i_t. That is, the transition matrix depends on X_t, x_t, and i_t,

$$P = P(X_t, x_t, i_t) \equiv [P_{jk}(X_t, x_t, i_t)].$$

Such dependence would ruin the tractability of our NL case (and thus make more difficult the AOP case as well), which exploited the exogeneity of the modes. However, as we need to go for numerical solutions in the BOP case anyway, these further complications are of less consequence.

Let

$$\tilde{P}(X_t, z_t, i_t, j_t, \varepsilon_t) \equiv P[X_t, \tilde{x}(X_t, z_t, i_t, j_t, \varepsilon_t), i_t],$$

where we have used (3.3). Then equation (4.6) is replaced by

$$p_{t+1|t} = \tilde{P}(X_t, z_t, i_t, j_t, \varepsilon_t)' p^+_{t|t}, \tag{5.6}$$

and (5.6) is used instead of (4.6) in the definition of $Q(s_t, z_t, i_t, j_t, \varepsilon_t, j_{t+1}, \varepsilon_{t+1})$. Furthermore, everywhere, P_{jk} is replaced by $\tilde{P}_{jk}(X_t, z_t, i_t, j_t, \varepsilon_t)$. The rest of the problems remains the same. Thus, formally, the extension to endogenous mode transitions is easy. However, as we have noted, the simplicity of the NL and AOP cases vanishes.

6 Examples

In this section we present some simple examples which help to illuminate the benefits of learning and experimentation. First we consider a backward-looking case, then add forward-looking components.

6.1 A backward-looking example

6.1.1 The model and benchmark specification

We consider the simplest possible example, where $n_X = 1$, $n_x = 0$, $n_i = 1$, $n_\varepsilon = 1$, and $N_j = \{1, 2\}$,

$$X_{t+1} = A_{j_{t+1}} X_t + B_{j_{t+1}} i_t + C_{j_{t+1}} \varepsilon_{t+1},$$

where ε_t is normally distributed with zero mean and unit variance. In our benchmark specification, we assume that $A_1 = A_2 = 1$ and $C_1 = C_2 = 1$, so

$$X_{t+1} = X_t + B_{j_{t+1}} i_t + \varepsilon_{t+1}.$$

Furthermore, $B_1 = -1.5$ and $B_2 = -0.5$. That is, the instrument i_t has a larger negative effect on X_{t+1} in mode 1 than in mode 2. We assume that the modes are quite persistent,

$$P \equiv \left[\begin{array}{cc} P_{11} & 1-P_{11} \\ 1-P_{22} & P_{22} \end{array} \right] = \left[\begin{array}{cc} 0.98 & 0.02 \\ 0.02 & 0.98 \end{array} \right].$$

Below we consider some sensitivity analysis to see how the results vary as we vary the different parameters. We also consider briefly the case where the uncertainty is over the state persistence coefficient A rather than the response to the instrument B.

It follows that the stationary distribution of the modes satisfies $\bar{p} \equiv (\bar{p}_1, \bar{p}_2)' = (0.5, 0.5)'$. We note that the predicted probability of mode 1 in period $t+1$, $p_{1,t+1|t}$, is similar to the perceived probability of mode 1 in period t, since the modes are so persistent,

$$p_{1,t+1|t} = p_{1t|t} P_{11} + (1 - p_{1t|t})(1 - P_{22}) = 0.02 + 0.96 p_{1t|t}. \tag{6.1}$$

We finally assume that the period loss function satisfies

$$L_t = \frac{1}{2} X_t^2.$$

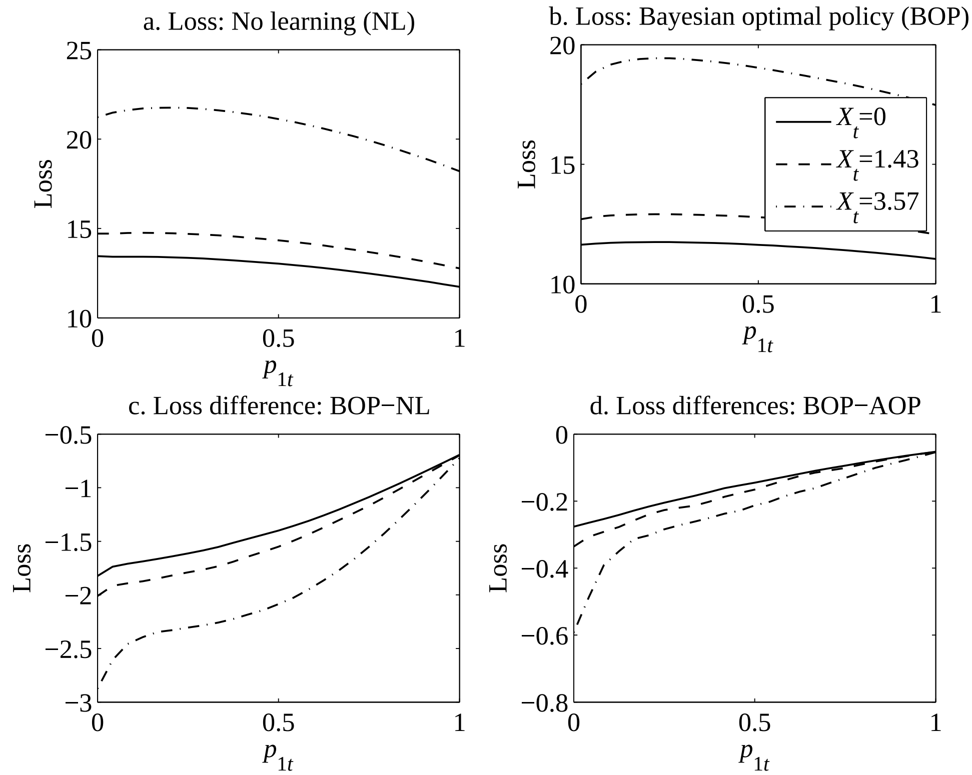
For this simple example, the state $s_t \equiv (X_t', p_{t|t}')'$ can be represented by $(X_t, p_{1t})'$, where we write p_{1t} for $p_{1,t|t}$, the perceived probability of mode 1 in period t.

Figure 6.1, panel a, shows the resulting value function $V(X_t, p_{1t})$ for the optimal policy under no learning (NL), as a function of p_{1t} for three different values of X_t.[16] Panel b shows the value function for the Bayesian optimal policy (BOP) as a function of p_{1t}, for the same three different values of X_t. Panel c plots the difference between the loss under BOP and NL. We see that the loss under BOP is significantly lower than under NL, albeit less so for high values of p_{1t}. Panel d shows the difference between the loss under BOP and the adaptive optimal policy (AOP). We see that the loss under BOP is lower than under AOP, but only modestly so.

Taken together, these results show that there is indeed benefit from learning in this example, although the additional benefits from experimentation are quite modest here. By moving from the NL case to AOP, and thus updating beliefs, policymakers are able to capture most of the benefit

[16] The example is solved with collocation methods via modifications of some of the programs of the CompEcon Toolbox described by Miranda and Fackler [12]

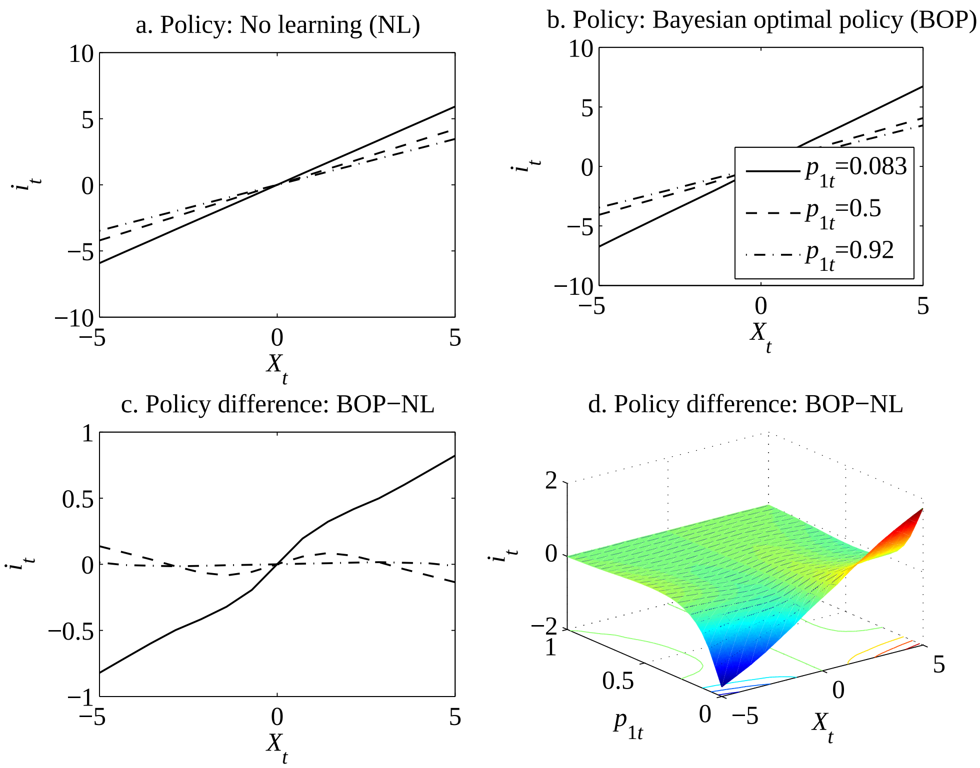
Figure 6.1: Losses from no learning (NL), adaptive optimal policy (AOP), and Bayesian optimal policy (BOP)



of the fully Bayesian optimal policy. The additional incremental improvement from AOP to BOP, arising from the experimentation motive, is much less significant. Thus, the AOP, which is relatively simple to compute and to implement recursively in real time, provides a good approximation to the fully optimal policy. Of course, these conclusions are dependent on the particular parameters chosen for this simple example, but as we show below we have found similar qualitative results in a number of other examples that we have analyzed.

Figure 6.2 shows the corresponding optimal policy functions. Panel a shows the optimal policy under NL as a function of X_t for three different values of p_{1t}. For given p_{1t}, the optimal policy function under NL is linear in X_t. Panel b shows the optimal policy function under BOP. On this scale, the nonlinearity in X_t for given p_{1t} is not apparent. Panel c shows the difference between the optimal policy under BOP and NL. Here we see that the Bayesian optimal policy is indeed nonlinear in X_t for given p_{1t}. Panel d plots the difference in the policies for all p_{1t} and all X_t in the interval $[-5, 5]$. We see that the difference is largest for small p_{1t}, where the Bayesian optimal policy

Figure 6.2: Policy for no learning (NL) and Bayesian optimal policy (BOP)

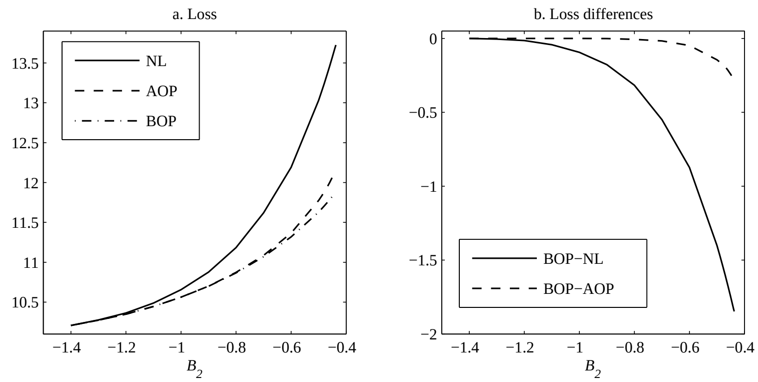


responds more aggressively (i_t is larger for positive values of X_t and smaller for negative values) than the adaptive policy. We discuss below how more aggressive policies can sharpen inference, and thus lessen future expected losses. But first we see how our results vary with the parameters of the model.

6.1.2 Sensitivity analysis

In order to better determine what drives the gains from learning and experimentation, we now consider some variations on the benchmark specification of the model. In each of the cases to follow we will report the losses and loss differences at a particular point in the state space, namely the center of it where $X_t = 0$ and $p_{1,t|t} = 0.5$. As we have seen above, the differences may be larger at other points in the space, but this provides a natural reference point and a convenient way to summarize each loss function by a single number. Further, in what follows we change each parameter one at a time, and thus we do not capture any potential interaction effects between the different parts of the model.

Figure 6.3: Losses and loss differences from no learning (NL), adaptive optimal policy (AOP), and Bayesian optimal policy (BOP) when the coefficient B_2 varies.



For our first experiment, we see how the gains from learning and experimentation vary as we vary the degree of differences across the modes. Thus, we fix $B_1 = -1.5$ and vary B_2, with our results shown in figure 6.3. As expected, when B_1 and B_2 are close together, there is little difference among the losses in the NL, AOP, and BOP cases. But as the difference increases, here by making B_2 smaller in magnitude, the gains from learning increase sharply, as the gap between the NL and AOP cases becomes visible in panel a. However, the gains from experimentation remain negligible until the difference in the coefficients across modes becomes more extreme. This is clear in panel b, where we see that the composite gains from learning and experimentation (BOP minus NL) are mostly due to learning, as the gap between BOP and AOP is near zero except at the rightmost edge of the figure. Thus, learning is beneficial even for moderate differences across modes, while experimentation only has noticeable benefits for more extreme uncertainty. But even in this more extreme range, the gains from experimentation are significantly smaller than the gains from learning.

In the next group of experiments, we keep the response coefficients fixed at $B_1 = -1.5$ and $B_2 = -0.5$ as in our benchmark specification, but vary other aspects of the model. First, we change the state persistence coefficient $A_1 = A_2 = A$ which was fixed at 1 in our benchmark specification. Our results in figure 6.4 show that when A is small, so that the state has less exogenous persistence, the gains from learning and experimentation are low. This is to be expected, as the gains from updating beliefs relate are due to improved knowledge which will help to make policy more accurate in the future. But if there is little persistence in the state, then actions today have little consequence

Figure 6.4: Losses and loss differences from no learning (NL), adaptive optimal policy (AOP), and Bayesian optimal policy (BOP) when $A_1 = A_2 = A$ varies (and B varies across modes).

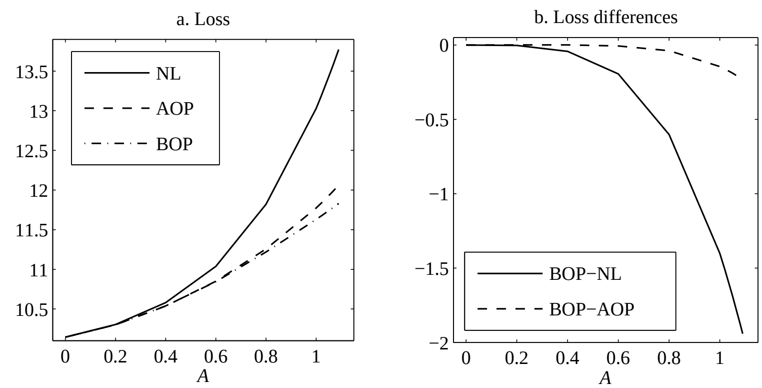
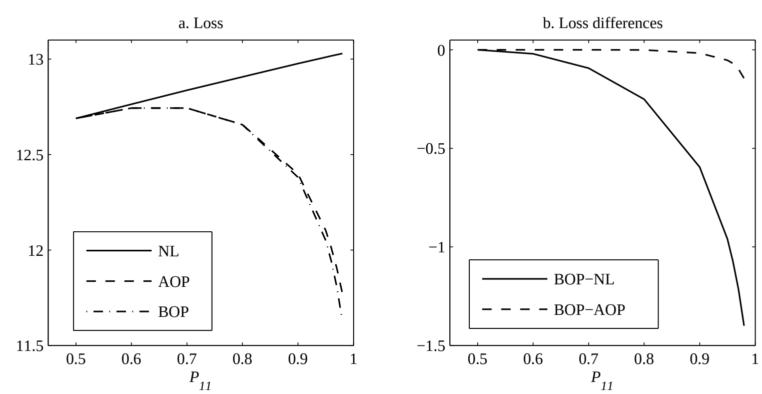


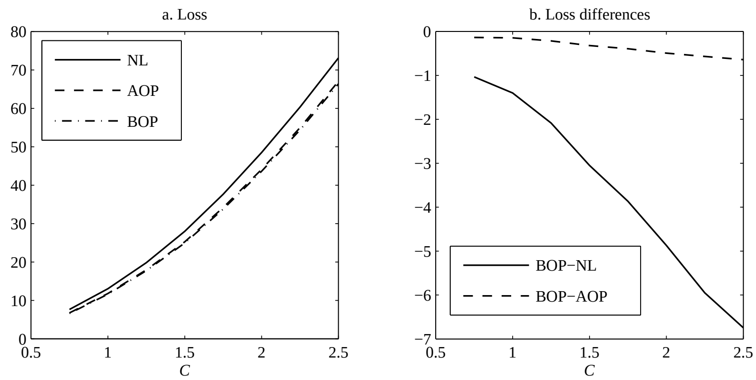
Figure 6.5: Losses and loss differences from no learning (NL), adaptive optimal policy (AOP), and Bayesian optimal policy (BOP) when $P_{11} = P_{22}$ varies (and B varies across modes).



for the future, and thus more accurate beliefs bring little or no gains. However, as the state becomes more persistent, the gains from learning increase substantially. The gains from experimentation become noticeable as well once the state becomes sufficiently persistent, but once again they are dwarfed by the gains from learning.

Along similar lines, we next consider the effects of the persistence of modes. We keep the transition matrix diagonal and symmetric, and vary the persistence as measured by $P_{11} = P_{22}$ which is set at 0.98 in the benchmark specification. Not surprisingly, when the modes are i.i.d. or nearly so, and thus P_{11} is near 0.5, there are no gains from learning or experimentation. Similarly to the previous case, there is no gain from increased accuracy of beliefs about the current mode if

Figure 6.6: Losses and loss differences from no learning (NL), adaptive optimal policy (AOP), and Bayesian optimal policy (BOP) when $C_1 = C_2 = C$ varies (and B varies across modes).

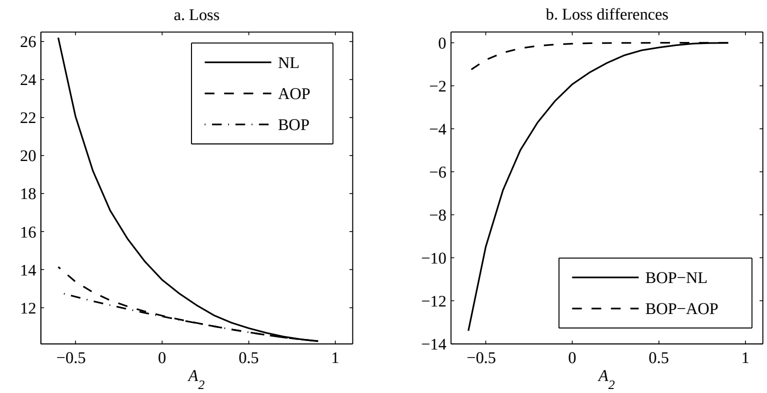


this knowledge has no consequence, which is the case here as the mode is about as likely to change as it is to stay the same. As the modes become more persistent, the gains from learning increase substantially while again the gains from experimentation are only sizeable at the very edge of the range considered when modes are nearly permanent. But once again the experimentation gains are much smaller in magnitude than the gains from learning, even when the modes are very persistent. One additional interesting aspect of this case is that as the modes become more persistent the level of the loss in the NL case increases, while it falls in the AOP and BOP cases. By exploiting the better knowledge of modes which comes with updating beliefs, in these cases policymakers are better able to tailor policy to the prevailing mode.

The final parameter variation we consider is the amount of volatility in the model, as measured by $C_1 = C_2 = C$. In our benchmark specification this is fixed at 1, but we now see what happens when the model is more or less variable. Not surprisingly, as we increase the variability, losses increase substantially in all cases. Moreover, the increases are nearly proportional in all the three cases. In absolute terms, the gains from learning increase sizeably as the variability increases, although by a much smaller amount as a percentage of the overall loss. Similarly, the absolute gains from experimentation are larger with more variability, but again much less so as a proportion of the total loss.

All of the above experiments considered the case where the policy response coefficient B varied across modes, which is perhaps the most natural starting case. However, now we analyze situations in which B is constant across modes (at -1.5) but the state persistence coefficient A varies. In

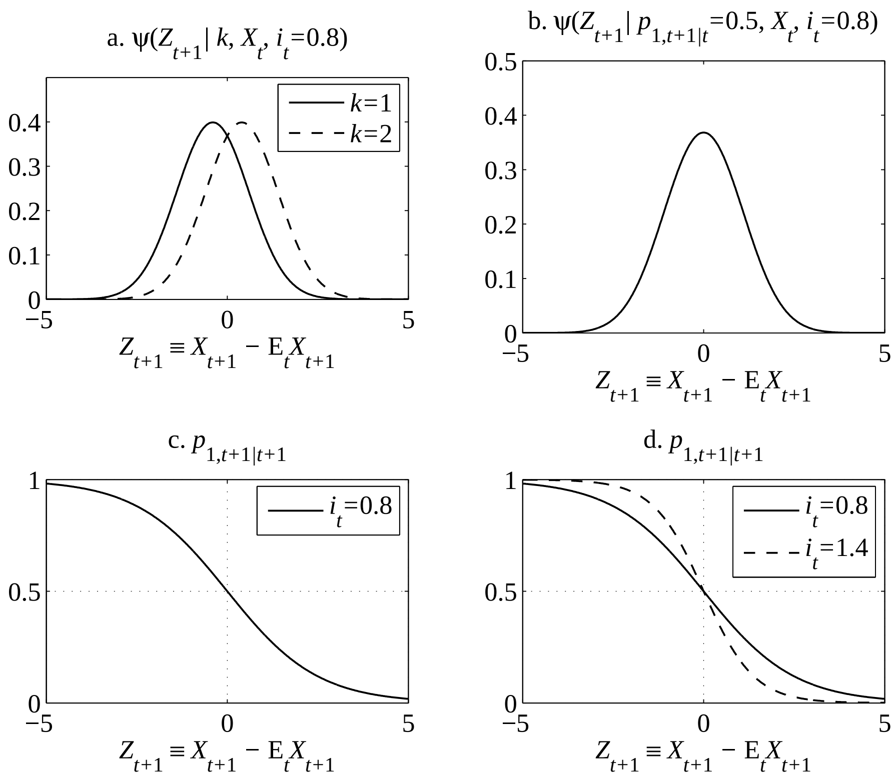
Figure 6.7: Losses and loss differences from no learning (NL), adaptive optimal policy (AOP), and Bayesian optimal policy (BOP) when A_2 varies (and differs from $A_1 = 1$).



particular, we fix $A_1 = 1$ and see what results when A_2 takes on different values. (The rest of the parameters are the same as in the benchmark specification.) Figure 6.7 summarizes our results, which in qualitative terms are quite similar to our previous findings. When A_2 is close to A_1, of course there is once again little gain from having sharper inference about modes. As the difference between A_1 and A_2 increases (as A_2 falls) the gains from learning grow substantially. Once again, only when the differences across modes are quite extreme are there sizeable experimentation gains, which yet again are significantly smaller than the gains from learning.

In summary, in all of the cases we have considered, the gains from learning and the gains from experimentation both increase in the situations where one would expect sharper inference to be beneficial. As the differences across modes grow larger, in the benchmark case where the instrument response coefficient B varies across modes as well as the case where the state persistence coefficient A varies, then the optimal policies conditional on each mode grow more different, and thus there are more gains from correctly inferring the current mode. Similarly, the gains from sharper inference increase as the effects of the current mode are longer lasting, either due to increased persistence in the state through A or more directly through increased persistence of the modes themselves through the transition matrix P. In all these cases the gains from learning are significant even for relatively modest amounts of uncertainty, while the gains from experimentation are only noticeable for more extreme uncertainty.

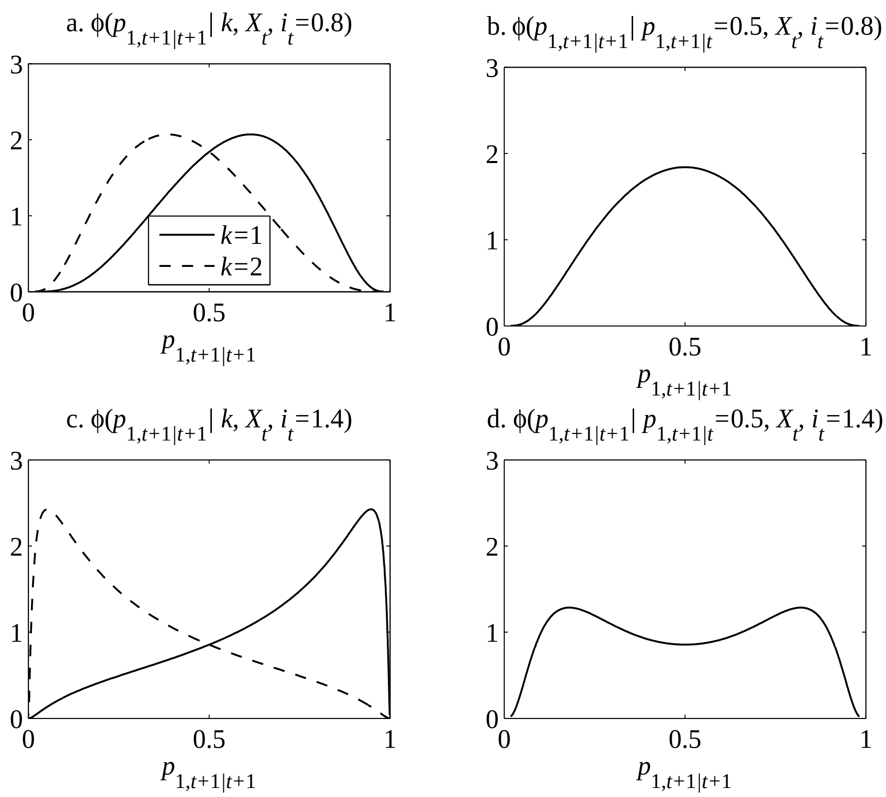
Figure 6.8: Probability density of X_{t+1} and updating of $p_{t+1|t+1}$



6.2 The interaction of learning and control

In order to better understand the nature of the different solutions and the role of learning, we consider figures 6.8 and 6.9 which depict how beliefs respond to different policies (all the calculations in this section use our benchmark specification). First, figure 6.8 shows the components of the Bayesian updating rule. Panel a shows the conditional density function of the innovation in X_{t+1}, $Z_{t+1} \equiv X_{t+1} - \mathrm{E}_t X_{t+1}$, conditional on the mode $j_{t+1} \equiv k$ where $k = 1$ or 2 in period $t+1$, for given X_t and i_t. Here X_t is set equal to 1, and i_t is set equal to 0.8; this value for i_t is approximately the optimal policy under NL for $X_t = 1$ and $p_{1,t+1|t} = p_{1t|t} = \bar{p}_1 = 0.5$. Panel b shows the *un*conditional (that is, not conditional on k) density function of the innovation in X_{t+1}, for $X_t = 1$, $i_t = 0.8$, and $p_{1,t+1|t} = 0.5$. Panel c plots the resulting updated $p_{1,t+1|t+1}$ as a function of the innovation in X_{t+1}. By Bayes Theorem, it is given by the ratio of the density of the innovation conditional on $k = 1$ to the unconditional density of the innovation multiplied by the period-t prediction of mode 1 in

Figure 6.9: Probability density of $p_{t+1|t+1}$



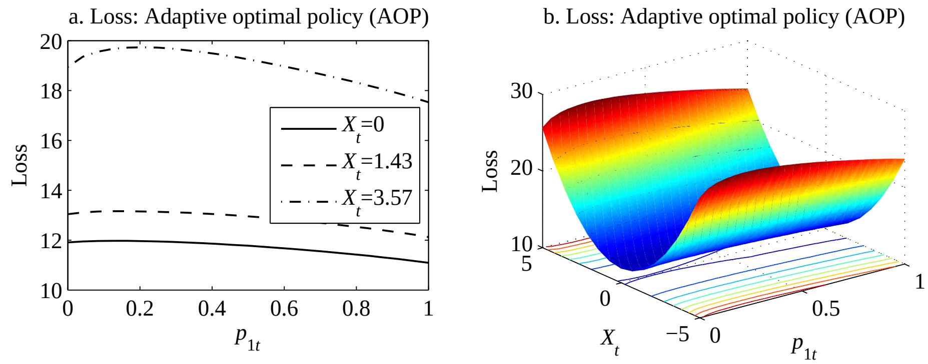
period $t+1$, $p_{1,t+1|t} = 0.5$,

$$p_{1,t+1|t+1} = \frac{\psi(X_{t+1} - \mathrm{E}_t X_{t+1} \,|\, k=1, X_t, i_t)}{\psi(X_{t+1} - \mathrm{E}_t X_{t+1} \,|\, p_{1,t+1|t}, X_t, i_t)}\, p_{1,t+1|t}. \tag{6.2}$$

We see that $p_{1,t+1|t+1}$ is decreasing in $X_{t+1} - \mathrm{E}_t X_{t+1}$. The larger the innovation in X_{t+1}, the less likely the mode 1, since, for a given positive i_t, mode 1 is associated with a larger negative effect of i_t on X_{t+1} and hence, everything else equal, a lower X_{t+1}. This is apparent in panel a, where the probability density of the innovation conditional on mode 1 is to the left of the density of the innovation conditional on mode 2.

Suppose now that the policymaker increases the value of the policy instrument, say from 0.8 to 1.4. Then, a larger value of the policy instrument multiplies the mode-dependent coefficient $B_{j_{t+1}}$. As a result, the conditional probability densities in panel a move further apart, and the unconditional density in panel b becomes more spread out. As a result, the updated $p_{1,t+1|t+1}$ becomes more sensitive to the innovation. This is shown in panel d, where $p_{1,t+1|t+1}$ as a function of the innovation is plotted for both $i_t = 0.8$ and $i_t = 1.4$. Thus, with a larger absolute value of the

Figure 6.10: Loss from adaptive optimal policy (AOP)



instrument, for a given realization of the innovation, the updated $p_{1,t+1|t+1}$ is closer to the extremes of 0 or 1. The policymaker becomes less uncertain about the mode in period $t+1$. In this sense, we can say that a larger instrument setting improves the updating and learning of the distribution of the modes. Thus, if the policymaker perceives that learning is beneficial, he or she would in this example be inclined to experiment by pursuing more aggressive policy, in the sense of increasing the magnitude of the instrument for a given X_t.

We will return shortly to the issue of when learning and experimentation is beneficial. But first, we note that, given the conditional and unconditional distribution of the innovation in X_{t+1} illustrated in figure 6.8, panels a and b, and the relation between the updated probability $p_{1,t+1|t+1}$ and the realization of the innovation in X_{t+1} illustrated in panel c, we can infer the conditional and unconditional probability densities of $p_{1,t+1|t+1}$.[17] These are shown in figure 6.9, panels a and b, respectively, for $i_t = 0.8$. Furthermore, panels c and d show the conditional and unconditional probability densities of $p_{1,t+1|t+1}$ when i_t is increased to 1.4. Comparing panels c and a, we see that a higher absolute value of the instrument moves the conditional densities of beliefs further apart. Thus, with a more aggressive policy, beliefs are much more sharply concentrated around the truth. Comparing panels d and b, we see that the unconditional density is further spread out, and in this case becomes bimodal. Thus, the mass of the unconditional distribution is closer to the extremes, 0 and 1, indicating that the uncertainty about the mode in period $t+1$ falls.

[17] If $\psi_p(p)$ and $\psi_Z(Z)$ denote the probability densities of scalars p and Z, and p is an invertible and continuously differentiable function of Z, $p = Q(Z)$, the densities are related by

$$\psi_p(p) = \psi_Z(Q^{-1}(p))dQ^{-1}(p)/dp.$$

When is learning beneficial? In order to understand this, we again look at figure 6.1, panel a, which shows the value function under NL, as a function of $p_{1,t|t}$ for three different values of X_t. Consider a policymaker in period t, with the perceived probability of mode 1 in period t equal to 0.5, so $p_{1,t|t} = 0.5$. Since 0.5 is the stationary probability for this Markov chain, this also means that the period-t predicted probability of mode 1 in period $t+1$, given by (6.1), is also 0.5. Under NL, the policymaker's predicted and updated probabilities are the same, $p_{1,t+1|t+1} = p_{1,t+1|t}$. Thus, in this case the conditional and unconditional probability distributions of $p_{1,t+1|t+1}$ in figure 6.9, panels a and b, are the same and are simply given by a spike with unit probability mass for $p_{1,t+1|t+1} = 0.5$.

Under adaptive optimal policy (AOP), the policymaker applies the same policy function as under NL, but now he uses Bayes Theorem to update the perceived probability of mode 1, $p_{1,t+1|t+1}$, after observing the innovation in X_{t+1} at the beginning of period $t+1$. That is, from the vantage point of period t, the updated probability $p_{1,t+1|t+1}$ in period $t+1$ is a random variable with the probability density shown in figure 6.9, panel b. As discussed above, the mean of this probability density is the predicted probability, $p_{1,t+1|t} = 0.5$. Comparing the perceived probability distribution of $p_{1,t+1|t+1}$ under AOP with what prevails under NL, we see a dramatic mean-preserving spread, from a spike with unit mass at 0.5 to the spread-out probability density shown in panel b.

As discussed above, such a mean-preserving spread reduces the intertemporal loss if the value function under NL is strictly concave as function of $p_{1,t+1|t+1}$. In this case Jensen's inequality implies that the expected future loss falls when the future beliefs become more dispersed. In figure 6.1, panel a, we see that the value function under NL indeed is concave, more so for higher values of X_{t+1} and lower values of $p_{1,t+1|t+1}$, but also in the vicinity of $p_{1,t+1|t+1} = 0.5$. Thus, we understand why the loss is lower under AOP, where the policymaker follows the same policy function, $i_{t+1} = F(X_{t+1}, p_{1,t+1|t+1})$, as under NL but updates the probability of mode 1 according to (6.2).

Under AOP, the policymaker does not consider adjusting the policy in order to change the shape of the density of $p_{1,t+1|t+1}$ and thereby improve the future precision of beliefs. Our previous discussion of figure 6.9 has revealed that increasing the absolute value of the instrument in this example will lead to a larger mean-preserving spread. In the case of increasing the instrument from 0.8 to 1.4, this increases the spread from that of the density in panel b to the that of the density in panel d. The value function under AOP is shown in figure 6.10. Compared with the value function under NL in panel a of figure 6.1, it is more concave for low values of $p_{1,t|t}$ and somewhat flatter for higher values.

Now, in the BOP case, the policymaker considers the influence of his policy on inference. Thus, he has the option of increasing the magnitude of the policy instrument somewhat, in order to increase the mean-preserving spread of the density of $p_{1,t+1|t+1}$, the benefit of which depends on the concavity of the AOP value function. The cost of this is an increase in the expected period loss in period $t+1$ from its minimum. The result of the optimal tradeoff is shown in panels c and d of figure 6.2 above. In this particular example, the policymaker chooses not to deviate much from the policy under NL and AOP. That is, he does not experiment much, except for small values of $p_{1,t+1|t} \approx p_{1t|t}$ where incidentally the concavity of the value function under AOP is the largest.[18] Furthermore, from figure 6.1, panels c and d, we see that the fall in the intertemporal loss from AOP to BOP is quite modest, and most of the fall in the loss arises in moving from NL to AOP.

Thus, in this example, the main benefit from learning arises without any experimentation. Although the amount of experimentation, measured as the policy difference between BOP and AOP, is substantial for low values of $p_{1t|t}$, the benefit in terms of additional loss is quite small. As we have seen in our sensitivity analysis, this holds true for most of the other parameterizations as well, and only when there are extreme differences across modes are the gains from experimentation sizeable. Furthermore, in the above example there is no direct cost whatsoever of a large instrument or a large change in the instrument. If such a cost is added, the magnitude and the benefits of experimentation (moving from AOP to BOP) shrink, whereas there is still substantial benefits from learning (moving from NL to AOP).

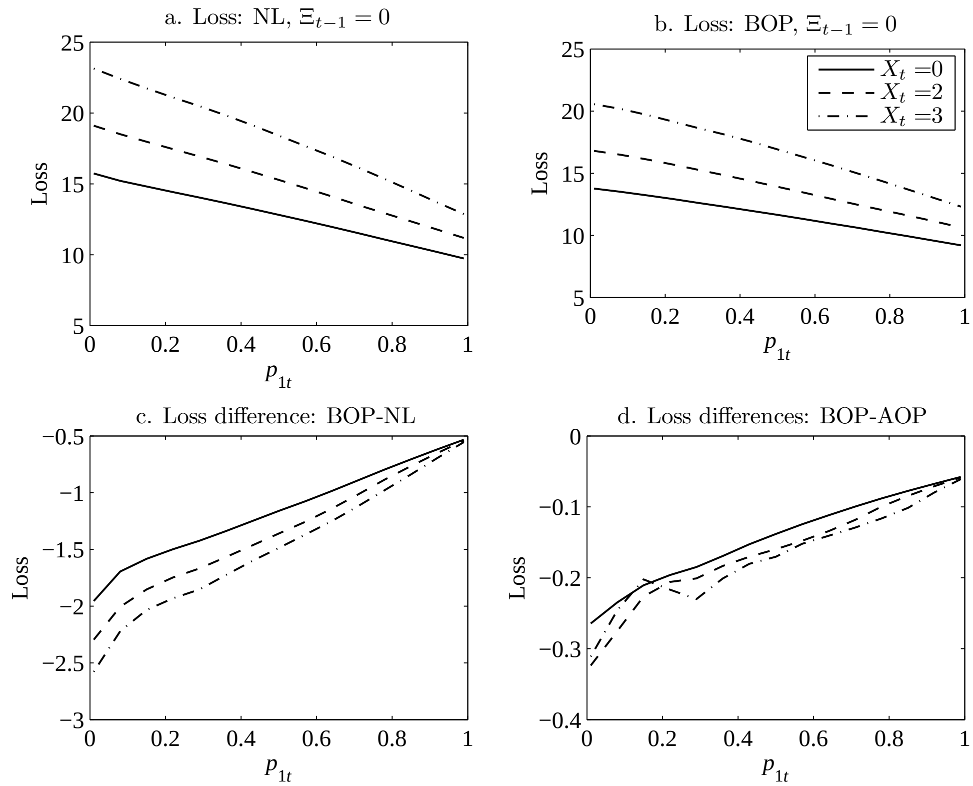
6.3 A forward-looking example

We now turn to a closely related example with forward-looking elements. The main implications of the backward-looking example are preserved, with one important qualification. The Lagrange multiplier associated with the equation for the forward-looking variable becomes a state variable, and this introduces some changes in the optimal policy in response to movements in this new multiplier state Ξ_{t-1}. Rather than being symmetric around $X_t = 0$, in the forward-looking case policies become asymmetric when $\Xi_{t-1} \neq 0$.

The example here is perhaps the simplest possible in the forward-looking case. There is one predetermined variable, one forward-looking variable, and two modes. The transition equation for

[18] The approximation $p_{1,t+1|t} \approx p_{1t|t}$ is justified by (6.1). Because the modes are so persistent, the predicted probability is close to the current perceived probability.

Figure 6.11: Losses from no learning (NL), adaptive optimal policy (AOP), and Bayesian optimal policy (BOP) for the forward-looking example with $\Xi_{t-1} = 0$.



the predetermined variable and the equation for the forward-looking variable are:

$$X_{t+1} = B_{j_{t+1}} i_t + \varepsilon_{t+1}, \tag{6.3}$$

$$\mathrm{E}_t x_{t+1} = X_t + x_t. \tag{6.4}$$

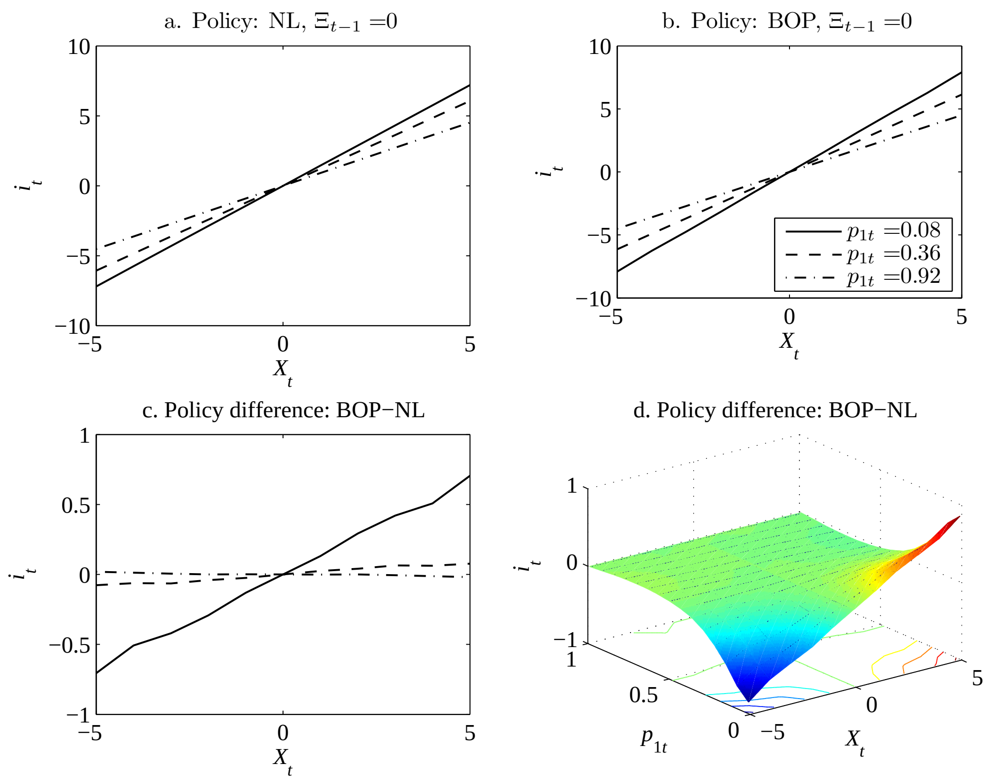
In the backward-looking example above, the uncontrolled system was a random walk which policy stabilized. The current system is similar, in that the jump variable x_t is essentially a random walk in the absence of control. As in the backward-looking case, we suppose that the instrument is more effective in mode 1:

$$B_1 = -1.5, \quad B_2 = -0.5.$$

Again, we assume that the modes are highly persistent with transition matrix:

$$P = \begin{bmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{bmatrix}.$$

Figure 6.12: Policy for no learning (NL) and Bayesian optimal policy (BOP) for the forward-looking example with $\Xi_{t-1} = 0$.



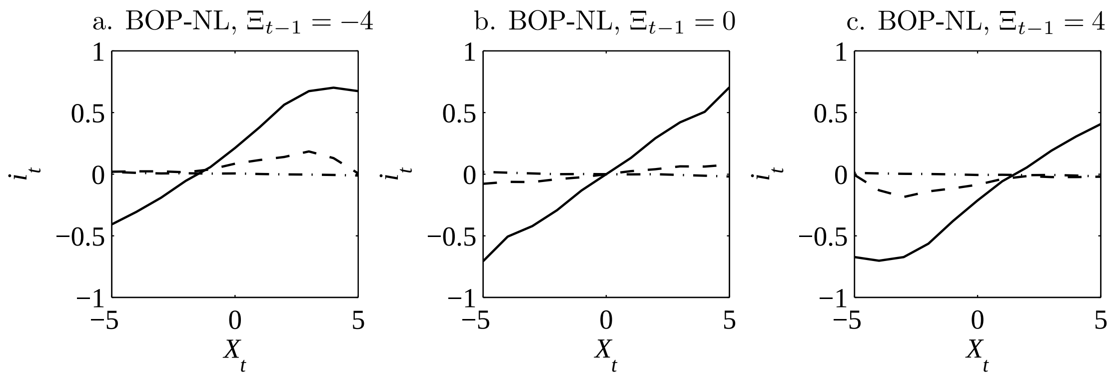
The loss function is similar in spirit to the backward-looking case, although different in details:

$$L_t = \frac{1}{2}x_t^2 + 0.1i_t^2, \quad \delta = 0.95.$$

Since the forward-looking variable x_t now has the random walk elements, it is also the one which receives the most weight in the loss function. We also include a small control-cost term.

Figure 6.11 is analogous to figure 6.1 above. In the current figure, panel a shows the resulting value function $V(X_t, \Xi_{t-1}, p_{1t})$ for the optimal policy under NL, as a function of p_{1t} for three different values of X_t, and with $\Xi_{t-1} = 0$. The shadow cost of the forward-looking constraint is zero, and thus this value is most comparable to the backward-looking case. Below we discuss the differences in results when the multiplier Ξ_{t-1} differs from zero. Again, panel b shows the value function for the BOP as a function of p_{1t}, while panel c plots the difference between the loss under BOP and NL, and panel d shows the difference between the loss under BOP and AOP. Overall, these results are quite similar to the backward-looking case. The value functions appear nearly linear with some modest concavity, suggesting that learning is beneficial but experimentation has

Figure 6.13: Differences in policy between Bayesian optimal policy (BOP) and no learning (NL) for the forward-looking example with different Ξ_{t-1} values. Solid line: $p_{1t} = 0.08$; dashed line: $p_{1t} = 0.36$; dot-dash line: $p_{1t} = 0.92$.



modest effects. Indeed, we again see that the loss under BOP is significantly lower than under NL, while the loss under BOP is lower than under AOP, but only modestly so.

Figure 6.12 is analogous to figure 6.2 above, showing the corresponding optimal policy functions. For the current figure, we again set $\Xi_{t-1} = 0$. Panel a shows the optimal policy under NL as a function of X_t for three different values of p_{1t}, while panel b shows the optimal policy function under BOP. As above, the nonlinearity in the BOP policy is not apparent at this scale. Panel c shows the difference between the optimal policy under BOP and NL, while panel d plots the difference in the policies for all p_{1t} and all X_t in the interval $[-5, 5]$. As in the backward-looking case, the difference between policies is largest for small p_{1t}, where the Bayesian optimal policy responds more aggressively. Also note that, for a given p_{1t}, the magnitudes of the differences, and hence the effects of experimentation on policy, are symmetric about the X_t origin. That is, i_t is larger for positive values of X_t and smaller for negative values, but the absolute value of the effect on i_t is the same when $|X_t|$ is the same.

We now examine the effects of the forward-looking constraint, as summarized by different values of Ξ_{t-1}. A nonzero Ξ_{t-1} correspond to a constraint from previous commitment and will therefore increase the loss compared to when Ξ_{t-1} is zero. However, a more interesting effect is on the experimentation component of policy. In particular, for different Ξ_{t-1} values some asymmetries in the policy appear. This is evident in figure 6.13, which plots the differences in the optimal policy under BOP and NL for three different values of p_{1t} in each panel, now for different Ξ_{t-1}. Panel b repeats panel c of figure 6.12 with $\Xi_{t-1} = 0$, while in panel a we set $\Xi_{t-1} = -4$ and in panel c we set $\Xi_{t-1} = 4$. We see that, in each case, the experimentation component of policy tends to lead toward more aggressive policy, but this effect is altered by the multiplier Ξ_{t-1}. Comparing panel a to panel b, we see that when $\Xi_{t-1} < 0$ the experimentation component is greater for positive values of X_t and smaller for negative values. The converse happens in panel c, as when $\Xi_{t-1} > 0$ the experimentation component is smaller for positive values.

These differences reflect a feature of the tradeoff between experimentation and control which is absent in the backward-looking case. Experimentation tends to push toward more aggressive policy to sharpen the inference about the modes. However, when $\Xi_{t-1} < 0$, the forward-looking constraint implies a larger loss penalty for more negative X_t and i_t, which dampens this effect. But, for positive X_t, the loss is smaller with $\Xi_{t-1} > 0$, which amplifies the effect.

More precisely, in this case the term (2.6) that must be added to intertemporal loss function to represent previous commitments is

$$\Xi_{t-1}\frac{1}{\delta}x_t.$$

A negative Ξ_{t-1} hence makes it desirable to increase x_t, everything else equal. By (6.3) and (6.4), x_t is determined by

$$x_t = -X_t + \mathrm{E}_t x_{t+1} = -X_t - \mathrm{E}_t \sum_{\tau=0}^{\infty} X_{t+1+\tau} = -X_t - \mathrm{E}_t \sum_{\tau=0}^{\infty} B_{j_{t+1+\tau}} i_{t+\tau},$$

where we assume that the sums converge. Since $B_j < 0$ for $j = 1$ and 2, increasing x_t means increasing i_t. Hence, for $\Xi_{t-1} < 0$ (> 0) and for each X_t and p_{1t}, under both NL and BOP the optimal i_t is higher (lower), and more so for BOP.

In economic terms, with forward-looking variables in the model, the key considerations are not just sharpening inference versus inducing more volatility, but also influencing the expectations of future variables. As we have seen the optimal policy embodies a rather intricate tradeoff amongst these factors. However, it remains the case that the gains from optimal experimentation are much smaller than the gains from learning.

7 Conclusions

In this paper we have presented a relatively general framework for analyzing model uncertainty and the interactions between learning and optimization. While this is a classic issue, very little to date has been done for systems with forward-looking variables, which are essential elements of modern models for policy analysis. Our specification is general enough to cover many practical cases of interest, but yet remains relatively tractable in implementation. This is definitely true for cases when decision makers do not learn from the data they observe (our no-learning case) or when they do learn but do not account for learning in optimization (our adaptive optimal policy case). In both of these cases, we have developed efficient algorithms for solving for the optimal policy which can handle relatively large models with multiple modes and many state variables. However, in the case of the Bayesian optimal policy, where the experimentation motive is taken into account, we must solve more complex numerical dynamic programming problems. Thus, we are haunted by the curse of dimensionality, forcing us to study relatively small and simple models.

Thus, an issue of much practical importance is the size of the experimentation component of policy, and the losses entailed by abstracting from it. While our results in this paper are far from comprehensive, they suggest that in practical settings the experimentation motive may not be a concern. The above and similar examples that we have considered indicate that the benefits of learning (moving from NL to AOP) may be substantial, whereas the benefits from experimentation (moving from AOP to BOP) are modest or even insignificant. If this preliminary finding stands up to scrutiny, experimentation in economic policy in general and monetary policy in particular may not be very beneficial, in which case there is little need to face the difficult ethical and other issues involved in conscious experimentation in economic policy. Furthermore, the AOP is much easier to compute and implement than the BOP. To have this truly be a robust implication, more simulations and cases need to be examined. In particular, it will be important in future work to see how these results are affected in more realistic and empirically relevant settings.

Appendix

A Details of the algorithm for the no-learning case

Here we provide more detail on the setup of the model in the no-learning case and adapt the algorithms in Svensson and Williams [15] (DFT) to the specification with $C_{2j} \not\equiv 0$.

A.1 Setup

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

$$L_t = \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix}' \begin{bmatrix} Q_{11j} & Q_{12j} & N_{1j} \\ Q_{12j}' & Q_{22j} & N_{2j} \\ N_{1j}' & N_{2j}' & R_j \end{bmatrix} \begin{bmatrix} X_t \\ x_t \\ i_t \end{bmatrix}.$$

Then the dual loss is

$$\tilde{L}_t = L_t - \gamma_t' z_t + \Xi_{t-1}' \frac{1}{\delta} H_j x_t.$$

We now substitute in for x_t using

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

where in the last line we introduce new notation for the shock. Since we assume $C_{1j}\varepsilon_t$ is independent of $C_{2j}\varepsilon_t$, we find it useful to denote the shock ε_t in the forward-looking equation by v_t. After this substitution we want to express the laws of motion and dual loss in terms of the expanded state $\tilde{X}_t = [X_t', \Xi_{t-1}']'$ and the expanded controls $\tilde{\imath}_t = [z_t', i_t', \gamma_t']'$. Suppressing time and mode subscripts for the time being (all are t and j, respectively (except $t-1$ on Ξ_{t-1})), we see that the dual loss can be written explicitly as

$$\begin{aligned} \tilde{L}_t = {} & X' \left(Q_{11} + A_{xX}' Q_{22} A_{xX} + 2A_{xX}' Q_{12}' \right) X + 2X' \left(N_1 + Q_{12} A_{xi} + A_{xX}' Q_{22} A_{xi} + A_{xX}' N_2 \right) i \\ & + 2z' (A_{xz}' Q_{12}' + A_{xz}' Q_{22} A_{xX}) X + \Xi' \frac{1}{\delta} H A_{xX} X + \Xi' \frac{1}{\delta} H A_{xz} z + \Xi' \frac{1}{\delta} H A_{xi} i \\ & - \gamma' z + z' (A_{xz}' Q_{22} A_{xz}) z + i' (R + A_{xi}' Q_{22} A_{xi} + 2A_{xi}' N_2) i + 2z' (A_{xz}' N_2 + A_{xz}' Q_{22} A_{xi}) i \\ & + v' (A_{xv}' Q_{22} A_{xv}) v + \text{cross terms in } v, \end{aligned}$$

where we don't write out the cross terms since they have zero conditional expectations. Thus, we can write the dual loss (ignoring the cross terms in v)

$$\tilde{L}_t = \begin{bmatrix} \tilde{X}_t \\ \tilde{\imath}_t \end{bmatrix}' \begin{bmatrix} \tilde{Q}_j & \tilde{N}_j \\ \tilde{N}_j' & \tilde{R}_j \end{bmatrix} \begin{bmatrix} \tilde{X}_t \\ \tilde{\imath}_t \end{bmatrix} + v_t' \Lambda_j v_t,$$

where (again suppressing the j index)

$$\begin{aligned}
\tilde{Q} &= \begin{bmatrix} \tilde{Q}_{11} & \tilde{Q}_{12} \\ \tilde{Q}_{12}' & 0 \end{bmatrix}, \\
\tilde{Q}_{11} &= Q_{11} + A_{xX}' Q_{22} A_{xX} + 2A_{xX}' Q_{12}', \\
\tilde{Q}_{12} &= \frac{1}{2\delta} A_{xX}' H', \\
\tilde{N} &= \begin{bmatrix} \tilde{N}_{11} & \tilde{N}_{12} & 0 \\ \tilde{N}_{21} & \tilde{N}_{22} & 0 \end{bmatrix}, \\
\tilde{N}_{11} &= Q_{12} A_{xz} + A_{xX}' Q_{22} A_{xz}, \\
\tilde{N}_{12} &= N_1 + Q_{12} A_{xi} + A_{xX}' Q_{22} A_{xi} + A_{xX}' N_2, \\
\tilde{N}_{21} &= \frac{1}{2\delta} H A_{xz}, \\
\tilde{N}_{22} &= \frac{1}{2\delta} H A_{xi}, \\
\tilde{R} &= \begin{bmatrix} \tilde{R}_{11} & \tilde{R}_{12} & \tilde{R}_{13} \\ \tilde{R}_{12}' & \tilde{R}_{22} & 0 \\ \tilde{R}_{13}' & 0 & 0 \end{bmatrix}, \\
\tilde{R}_{11} &= A_{xz}' Q_{22} A_{xz}, \\
\tilde{R}_{12} &= A_{xz}' N_2 + A_{xx}' Q_{22} A_{xi}, \\
\tilde{R}_{13} &= -I/2, \\
\tilde{R}_{22} &= R + A_{xi}' Q_{22} A_{xi} + 2A_{xi}' N_2, \\
\Lambda &= A_{xv}' Q_{22} A_{xv}.
\end{aligned}$$

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

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

where

$$\tilde{\varepsilon}_{t+1} = \begin{bmatrix} \varepsilon_{t+1} \\ \nu_t \end{bmatrix}, \quad \tilde{A}_{jk} = \begin{bmatrix} A_{11k} + A_{12k} A_{xXj} & 0 \\ 0 & 0 \end{bmatrix},$$

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

Furthermore, for the case where $C_{2j} \equiv 0$ and the forward variables do not reveal the mode j, we have that A_{xX}, A_{xz}, A_{xi} are independent of the mode and $A_{xv} \equiv 0$, so the dependence on j in $\tilde{A}_{jk}$, $\tilde{B}_{jk}$, and $\tilde{C}_{jk}$ disappears.

The value function for the dual problem, $\tilde{V}(X_t, p_{t|t})$, 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 we 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 3.) They will satisfy the Bellman equation

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

where

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

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

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

We can rewrite the first-order conditions 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

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

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 $\hat{V}(p_t)_j$:

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

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

$$\hat{w}(p_t)_j = \mathrm{tr}(\Lambda_j) + \delta \sum_k P_{jk} [\mathrm{tr}(\hat{V}(P' p_t)_k \tilde{C}_{jk} \tilde{C}_{jk}') + \hat{w}(P' p_t)_k]. \tag{A.3}$$

The value function for the primal problem is

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

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

$$x_t = F_{x\tilde{X}}(p_t)_j \tilde{X}_t + F_{xv}(p_t)_j v_t.$$

We may also find the conditional value function

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

A.2 The algorithm

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 the appendix of DFT.

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

$$
\begin{aligned}
J^\tau &\equiv \sum_j p_{j,t+\tau,t} \left[\tilde{R}_j + \delta \sum_k P_{jk} \tilde{B}'_{jk} \hat{V}_k^{\tau+1} \tilde{B}_{jk} \right], \\
K^\tau &\equiv \sum_j p_{j,t+\tau,t} \left[\tilde{N}'_j + \delta \sum_k P_{jk} \tilde{B}'_{jk} \hat{V}_k^{\tau+1} \tilde{A}_{jk} \right], \\
\tilde{F}^\tau &= -\,(J^\tau)^{-1} K^\tau, \\
\hat{V}_j^\tau &= \tilde{Q}_j + \tilde{N}_j \tilde{F}^\tau + \tilde{F}^{\tau\prime} \tilde{N}'_j + \tilde{F}^{\tau\prime} \tilde{R}_j \tilde{F}^\tau \\
&\quad + \delta \sum_k P_{jk} [\tilde{A}'_{jk} + \tilde{F}^{\tau\prime} \tilde{B}'_{jk}] \hat{V}_k^{\tau+1} [\tilde{A}_{jk} + \tilde{B}_k \tilde{F}^\tau)], \\
\hat{w}_j^\tau &= \operatorname{tr}(\Lambda_j) + \delta \sum_k P_{jk} [\operatorname{tr}(\hat{V}_k^{\tau+1} \tilde{C}_{jk} \tilde{C}'_{jk}) + \hat{w}_k^{\tau+1}], \\
\tilde{V}^\tau &= \sum_j p_{j,t+\tau,t} \hat{V}_j^\tau, \\
w_j^\tau &= \sum_j p_{j,t+\tau,t} \hat{w}_j^\tau.
\end{aligned}
$$

This procedure 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 3) 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

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

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

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

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

B Verifying the law of iterated expectations in the case of Bayesian optimal policy

It will be slightly simpler to use the general probability measure notation, $\Pr(\cdot\,|\,\cdot)$, although we will translate this to the specific cases at the end. We also write p_t for $p_{t|t}$, for simplicity. Finally, for simplicity we only consider the case without forward-looking variables (so we need only deal with X_t rather than $\tilde{X}_t$). The generalization to forward-looking variables is straightforward.

Thus, we want to verify

$$\mathrm{E}_t \hat{V}(s_{t+1}, j_{t+1}) = \mathrm{E}_t V(s_{t+1}),$$

where $V(s_t) \equiv \mathrm{E}_t \hat{V}(s_t, j_t)$.

First, in the BOP case, we note that we can write $p_{t+1} = \hat{Q}(X_{t+1}; X_t, p_t, i_t)$, and so we can define

$$\breve{V}(X_{t+1}, j_{t+1}; X_t, p_t, i_t) \equiv \hat{V}(X_{t+1}, \hat{Q}(X_{t+1}; X_t, p_t, i_t), j_{t+1}).$$

Then we consider

$$\mathrm{E}_t \hat{V}(X_{t+1}, p_{t+1}, j_{t+1}) \equiv \int \breve{V}(X_{t+1}, j_{t+1}; X_t, p_t, i_t) d \Pr(X_{t+1}, j_{t+1} \,|\, \mathcal{X}_t), \tag{B.1}$$

where the identity specifies the notation for the joint probability measure of (X_{t+1}, j_{t+1}), $\Pr(X_{t+1}, j_{t+1} \,|\, \mathcal{X}_t)$, conditional on the information set in period t, $\mathcal{X}_t \equiv \sigma(\{X_t, X_{t-1}, ...\})$ (that is, the sigma-algebra generated by current and past realizations of X_s, $s \le t$). We note that $p_t = \mathrm{E}(j_t \,|\, \mathcal{X}_t)$ is $\mathcal{X}_t$-measurable, that is, p_t is a function of $\mathcal{X}_t$. Furthermore, i_t is $\mathcal{X}_t$-measurable. Hence, $\mathrm{E}_t[\cdot] \equiv \mathrm{E}[\cdot \,|\, \mathcal{X}_t, p_t, i_t] \equiv \mathrm{E}[\cdot \,|\, \mathcal{X}_t]$. Also, we note that we can write

$$\mathrm{E}_{t+1} \hat{V}(X_{t+1}, p_{t+1}, j_{t+1}) \equiv \int \breve{V}(X_{t+1}, j_{t+1}; X_t, p_t, i_t) d \Pr(j_{t+1} \,|\, \mathcal{X}_{t+1}) \equiv V(X_{t+1}, p_{t+1}).$$

We will use two equivalent decompositions of the joint measure. First, perhaps the most natural decomposition is

$$\begin{aligned}
\Pr(X_{t+1}, j_{t+1} = k \,|\, \mathcal{X}_t) &= \Pr(X_{t+1} \,|\, j_{t+1} = k, \mathcal{X}_t) \Pr(j_{t+1} = k \,|\, \mathcal{X}_t) \\
&= \sum_j \Pr(X_{t+1} \,|\, j_{t+1} = k, \mathcal{X}_t) \Pr(j_{t+1} = k \,|\, j_t = j) \Pr(j_t = j \,|\, \mathcal{X}_t) \\
&= \sum_j \Pr(X_{t+1} \,|\, j_{t+1} = k, \mathcal{X}_t) P_{jk} p_{jt}.
\end{aligned} \tag{B.2}$$

Alternatively, we can decompose the joint measure as

$$\Pr(X_{t+1}, j_{t+1} = \ell \,|\, \mathcal{X}_t) \;=\; \Pr(j_{t+1} = \ell \,|\, X_{t+1}, \mathcal{X}_t) \Pr(X_{t+1} \,|\, \mathcal{X}_t)$$

$$\begin{aligned}
&= \Pr(j_{t+1} = \ell \,|\, \mathcal{X}_{t+1}) \sum_j \Pr(X_{t+1} \,|\, j_t = j, \mathcal{X}_t) \Pr(j_t = j \,|\, \mathcal{X}_t) \\
&= \Pr(j_{t+1} = \ell \,|\, \mathcal{X}_{t+1}) \sum_{j,k} \Pr(X_{t+1} \,|\, j_t = j, j_{t+1} = k, \mathcal{X}_t) \Pr(j_{t+1} = k \,|\, j_t = j) \Pr(j_t = j \,|\, \mathcal{X}_t). \\
&= p_{\ell,t+1} \sum_{j,k} \Pr(A_k X_t + B_k i_t + C_k \varepsilon_{t+1} \,|\, j_t = j, j_{t+1} = k, \mathcal{X}_t) P_{jk} p_{jt} \\
&= p_{\ell,t+1} \sum_{j,k} \varphi(\varepsilon_{t+1}) P_{jk} p_{jt}. \qquad \text{(B.3)}
\end{aligned}$$

Thus, using the first decomposition, (B.2), with (B.1) we have an expression as in section 5.2,

$$\begin{aligned}
&\mathrm{E}_t \hat{V}(X_{t+1}, p_{t+1}, j_{t+1}) \\
&= \int \sum_{j,k} \breve{V}(A_k X_t + B_k i_t + C_k \varepsilon_{t+1}, k; X_t, p_t, i_t) P_{jk} p_{jt} \varphi(\varepsilon_{t+1}) d\varepsilon_{t+1} \\
&= \int \sum_{j,k} \hat{V}[A_k X_t + B_k i_t + C_k \varepsilon_{t+1}, Q(A_k X_t + B_k i_t + C_k \varepsilon_{t+1}; X_t, p_t), k] P_{jk} p_{jt} \varphi(\varepsilon_{t+1}) d\varepsilon_{t+1}
\end{aligned}$$

On the other hand, using the second decomposition, (B.3), we can write (B.1) as

$$\begin{aligned}
&\mathrm{E}_t \hat{V}(X_{t+1}, p_{t+1}, j_{t+1}) \\
&= \int \sum_{j,k,\ell} \breve{V}(X_{t+1}, \ell; X_t, p_t, i_t) p_{\ell,t+1} \; P_{jk} p_{jt} \varphi(\varepsilon_{t+1}) d\varepsilon_{t+1} \\
&= \int \sum_{j,k} V(X_{t+1}, \hat{Q}(X_{t+1}; X_t, p_t)) P_{jk} p_{jt} \varphi(\varepsilon_{t+1}) d\varepsilon_{t+1} \\
&= \int \sum_{j,k} V[A_k X_t + B_k i_t + C_k \varepsilon_{t+1}, Q(A_k X_t + B_k i_t + C_k \varepsilon_{t+1}; X_t, p_t)] P_{jk} p_{jt} \varphi(\varepsilon_{t+1}) d\varepsilon_{t+1} \\
&= \mathrm{E}_t V(X_{t+1}, p_{t+1})
\end{aligned}$$

Note that, by averaging with respect to p_t, we thus eliminate j_t as a state variable and do not need to compute the conditional value function $\hat{V}(X_t, p_t, j_t)$.

References

[1] Beck, Günter W., and Volker Wieland (2002), "Learning and Control in a Changing Economic Environment," *Journal of Economic Dynamics and Control* 25, 1359–1377.

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

[3] Cogley, Timothy, Riccardo Colacito, and Thomas J. Sargent (2007) "The Benefits from U.S. Monetary Policy Experimentation in the Days of Samuelson and Solow and Lucas," *Journal of Money Credit and Banking* 39, s67–99.

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

[5] 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.

[6] Ellison, Martin (2006), "The Learning Cost of Interest Rate Reversals," *Journal of Monetary Economics* 53, 1895–1907.

[7] Ellison, Martin, and Natacha Valla (2001), "Learning, Uncertainty and Central Bank Activism in an Economy with Strategic Interactions," *Journal of Monetary Economics* 48, 153–171.

[8] Evans, George, and Seppo Honkapohja (2001) *Learning and Expectations in Macroeconomics*, Princeton University Press, Princeton.

[9] Kiefer, Nicholas M. (1989), "A Value Function Arising in the Economics of Information," *Journal of Economic Dynamics and Control* 13, 201–223.

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

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

[12] Miranda, Mario J., and Paul L. Fackler (2002), *Applied Computational Economics and Finance*, MIT Press.

[13] 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.

[14] Svensson, Lars E.O. (2007), “Optimization under Commitment and Discretion, the Recursive Saddlepoint Method, and Targeting Rules and Instrument Rules: Lecture Notes,” www.princeton.edu/svensson.

[15] Svensson, Lars E.O., and Noah Williams (2007), “Monetary Policy with Model Uncertainty: Distribution Forecast Targeting,” working paper, www.princeton.edu/svensson/ and www.princeton.edu/∼noahw/.

[16] Tesfaselassie, Mewael F., Eric Schaling, and Sylvester C. W. Eijffinger (2006), “Learning about the Term Structure and Optimal Ruels for Inflation Targeting,” CEPR Discussion paper No. 5896.

[17] Wieland, Volker (2000a), “Learning by Doing and the Value of Optimal Experimentation,” *Journal of Economic Dynamics and Control* 24, 501–534.

[18] Wieland, Volker (2006), “Monetary Policy and Uncertainty about the Natural Unemployment Rate: Brainard-Style Conservatism versus Experimental Activism,” *Advances in Macroeconomics* 6(1).

[19] 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.