FEDERAL
RESERVE
BANK
of ATLANTA

Learning and Monetary Policy Shifts

Frank Schorfheide

Working Paper 2003-23
October 2003

Working Paper Series

Federal Reserve Bank of Atlanta
Working Paper 2003-23
October 2003

Learning and Monetary Policy Shifts

Frank Schorfheide, University of Pennsylvania

Abstract: This paper estimates a dynamic stochastic equilibrium model in which agents use a Bayesian rule to learn about the state of monetary policy. Monetary policy follows a nominal interest rate rule that is subject to regime shifts. The following results are obtained. First, the author's policy regime estimates are consistent with the popular view that policy was marked by a shift to a high-inflation regime in the early 1970s, which ended with Volcker's stabilization policy at the beginning of the 1980s. Second, while Bayesian posterior odds favor the "full-information" version of the model in which agents know the policy regime, the fall of inflation and interest rates in the disinflation episode in the early 1980s is better captured by the delayed response of the "learning" specification. Third, the author examines the magnitude of the expectations-formation effect of monetary policy interventions in the "learning" specification by comparing impulse responses to a version of the model in which agents ignore the information contained in current and past monetary policy shocks about the likelihood of a regime shift.

JEL classification: C11, C32, E52

Key words: Bayesian econometrics, learning, monetary DSGE models

---

The author would like to thank Thomas Lubik for helpful comments and suggestions. Financial support from the Research Foundation of the University of Pennsylvania is gratefully acknowledged. Gauss programs to implement the empirical analysis conducted in the paper will be available at http://www.econ.upenn.edu/~schorf. This paper was presented at the Monetary Policy and Learning Conference sponsored by the Federal Reserve Bank of Atlanta in March 2003. The views expressed here are the author's and not necessarily those of the Federal Reserve Bank of Atlanta or the Federal Reserve System. Any remaining errors are the author's responsibility.

Please address questions regarding content to Frank Schorfheide, Department of Economics, University of Pennsylvania, McNeil Building, 3718 Locust Walk, Philadelphia, Pennsylvania 19104-6297, 215-898-8486, 215-573-2057 (fax), schorf@ssc.upenn.edu

Learning and Monetary Policy Shifts

1 Introduction

The effects of monetary policy are closely linked to the expectation formation of agents in the economy. Vector autoregressions (VAR) in the tradition of Sims (1980) have emerged as a tool to assess the consequences of policy interventions. The intervention is modelled as an unanticipated deviation – a policy shock – from the perceived policy rule, and its consequences are evaluated through impulse response functions. However, extended systematic deviations from the perceived policy rule may lead agents to change their beliefs about the conduct of monetary policy and invalidate the VAR impulse response predictions.

One method to overcome this problem is to estimate a fully-specified dynamic stochastic general equilibrium (DSGE) model that can be re-solved for alternative policy rules to predict effects of fundamental changes in the policy regime. However, this approach faces some conceptual difficulties as well (e.g. Sims (1982) and Cooley, LeRoy, and Raymon (1984)). First, for several periods after the regime change the agents are potentially uncertain whether the policy shift was temporary or permanent and the transition dynamics are possibly affected by the learning process. Second, the extent to which past data can be used to validate the predictions is very limited, since the policy change typically has no precedent.

In this paper we estimate a basic New Keynesian monetary DSGE model, along the lines of King (2000), in which monetary policy follows a regime switching process. Agents do take the possibility of regime shifts into account when they translate observed monetary policy into expectations about future output, prices, and interest rates. Unlike in the model considered in Sargent (1999), our regime switching framework offers no explanation why monetary policy shifts occur over time. We simply assume that there are high-inflation and low-inflation regimes and that the transition probabilities stay constant. While firms and households learn about the conduct of monetary policy, the central bank itself does not attempt to learn about the effectiveness of their policy and choose policy in an optimal manner.

We use a Bayesian approach that combines a prior distribution with the likelihood function derived from the structural model to obtain posterior estimates. Our empirical analysis has three parts. First, together with the parameter estimates we generate estimates of the monetary policy regimes in the post-war U.S. Our estimates are consistent with the popular view that policy was marked by a shift to a high-inflation regime in the early 1970's which ended with Volcker's stabilization policy at the beginning of the 1980's.

Second, we study the empirical evidence in favor of the learning mechanism. As an alternative, we consider a version of the DSGE model in which agents have full information about the current state of monetary policy. While Bayesian posterior odds favor the 'full-information' version of the model, the fall of inflation and interest rates in the disinflation episode in the early 1980's is better captured by the delayed response of the 'learning' specification.

Third, we examine the magnitude of the expectations-formation effect of monetary policy interventions in the 'learning' specification by comparing impulse responses to a version of the model in which agents ignore the information contained in current and past monetary policy shocks about the likelihood of a regime shift.

The likelihood-based estimation of a DSGE model with policy-regime shifts in general leads to a very complicated non-linear filtering problem that would take a long time to solve on a conventional computer.[1] We make two simplifications. First, we use a log-linear approximation of the DSGE model, and second we assume that the policy rule depends only on observed variables, except for the regime indicator and the policy shock, and not on latent model variables such as potential output. We show that under these assumptions it is fairly straightforward to compute the likelihood function of the regime switching model, using a modification of the Kalman filter. Methods described in Schorfheide (2000) are employed to evaluate the posterior distributions.

Markov-switching models have been used by Sims (2000a) and Sims and Zha

[1]Substantial progress in this direction has been made by Fernandez-Villaverde and Rubio-Ramirez (2002), who implement the likelihood-based estimation of a stochastic growth model without using a log-linear approximation of the model solution.

(2002) to study monetary policy. Sims (2000a) considers a univariate policy reaction function that is subject to regime shifts, whereas Sims and Zha (2002) extend the analysis to an identified VAR. While both papers find frequent oscillation between regimes that indicate no fundamental shift of U.S. monetary policy in the post-war period, our estimation uncovers essentially two distinct shifts of monetary policy.

DSGE models with shifting policy regimes have recently been analyzed by Andolfatto, Moran, and Hendry (2002), Andolfatto and Gomme (2003), and Erceg and Levin (2001). The first two papers consider cash-in-advance models. Andolfatto, Moran, and Hendry (2002) show that the agents' learning can explain the failure of conventional tests of unbiasedness of inflation expectations. Andolfatto and Gomme (2003) use their model to study the Canadian disinflation episode. Our findings with respect to the effects of a disinflation policy in the 'learning' and 'full-information' environment, by and large, resemble the results reported by Andolfatto and Gomme (2003). Erceg and Levin (2001) demonstrate that the 'learning' mechanism is able to generate inflation persistence in a DSGE model with staggered nominal contracts. However, none of these papers formally estimates the full DSGE model.

Leeper and Zha (2002) examine the important question to what extent monetary policy interventions trigger significant expectation-formation effects. The authors consider a simple two-equation model with an exogenous money-growth rate rule that is subject to regime shifts and propose a measure of modesty of policy interventions. This measure can be computed from an identified VAR and, under certain assumptions, provides an indication of the magnitude of the expectation-formation effects. In the third part of our empirical analysis we directly study the magnitude of the expectation formation based on our estimated DSGE model. We find that even small interventions can trigger a substantial 'learning' effect and lead to different predictions than the 'full-information' version of the model. We offer two explanations. First, in a model with a feedback monetary policy rule 'learning' affects the contemporaneous impact of the policy shock. Second, small policy shocks are historically associated with the low inflation regime. Hence, Bayesian updating leads agents to interpret a modest intervention as evidence for the low inflation target and, vice versa, large interventions as evidence for the high inflation regime.

The paper is organized as follows. Section 2 presents the monetary DSGE model and its approximation. Section 3 describes the econometric approach and provides some details about the computation of the likelihood function. Empirical results are presented in Section 4 and Section 5 concludes.

2 The Model Economy

We consider a monetary business cycle model with optimizing households and monopolistically competitive firms that face price stickiness. The model is a variant of the so-called New Keynesian IS-LM model. The main difference is that the monetary policy rule is subject to regime shifts. To keep the paper self-contained we outline the structure of the model. Detailed derivations can be found, for instance, Galí and Gertler (1999), King (2000), King and Wolman (1999), and Woodford (2000). The main difference between the model presented below and the earlier work is that the policy rule is subject to regime shifts.

2.1 The Firms

The production sector is described by a continuum of monopolistically competitive firms each of which faces a downward-sloping demand curve for its differentiated product $j \in [0, 1]$:

$$p_t(j) = \left(\frac{y_t(j)}{Y_t} \right)^{-1/\nu} P_t. \tag{1}$$

This demand function can be derived in the usual way from Dixit-Stiglitz preferences. $y_t(j)$ and $p_t(j)$ are quantity and price of good j, whereas Y_t and

$$P_t = \left[\int_0^1 p_t(j)^{1-\nu} \right]^{1/(1-\nu)}$$

refer to the aggregate demand and price level, respectively: The parameter ν is the elasticity of substitution between two differentiated goods. As $\nu \to \infty$ the demand function becomes perfectly elastic and the differentiated goods become substitutes.

Production is assumed to be linear in labor $n_t(j)$, which each firm hires on a competitive labor market:

$$Y_t(j) = A_t n_t(j). \tag{2}$$

Total factor productivity is an exogenous process that follows a random walk with drift

$$\ln A_t = \gamma + \ln A_{t-1} + \widetilde{z}_t, \tag{3}$$

where $\widetilde{z}_t$ is an AR(1) process

$$\widetilde{z}_t = \rho_z \widetilde{z}_{t-1} + \epsilon_{z,t}.$$

The technology A_t induces a stochastic trend in the model economy. The autoregressive structure of $\widetilde{z}_t$ captures the observed serial correlation in output growth, as the benchmark model lacks a sophisticated internal propagation mechanism.

To introduce nominal rigidity, we follow Calvo (1983) and assume that each firm has a constant probability $1-\eta$ of being able to re-optimize its price $P_t(j)$. Whenever, firms are not able to re-optimize they adjust their price from the previous period by the steady state (gross) inflation rate π. The optimal price $P_t^o(j)$ is determined by maximizing

$$\mathbb{E}_t \left[\sum_{s=0}^{\infty} (\beta\eta)^s q_{t+s} \Pi_{t+s} \right], \quad \Pi_{t+s} = \left(\frac{P_t^o(j)\pi^s}{P_{t+s}} Y_{t+s}(j) - W_{t+s} n_{t+s}(j) \right) \tag{4}$$

subject to (1) and (2), where W_{t+s} is the wage, β is the fixed discount factor, and q_{t+s} is a (potentially) time-dependent discount factor that firms use to evaluate future profit streams. The firms take aggregate demand and price level in (1) as given. The aggregate price level is derived by noting that the fraction of firms that optimized its price s periods ago and charges the price $P_t^o(j)\pi^s$ is $(1-\eta)\eta^s$. Thus,

$$\begin{aligned} P_t &= \left[(1-\eta) \sum_{s=0}^{\infty} \eta^s (\pi^s P_{t-s}^o)^{1-\nu} \right]^{1/(1-\nu)} \qquad (5) \\ &= \left[(1-\eta)(P_t^o)^{1-\nu} + \eta(\pi P_{t-1})^{1-\nu} \right]^{1/(1-\nu)}. \end{aligned}$$

Define the inflation rates $\pi_t = P_t/P_{t-1}$ and $\pi_t^o = P_t^o/P_{t-1}$ and divide (5) by P_{t-1} to obtain the following law of motion for aggregate inflation:

$$\pi_t = \left[\eta\pi^{1-\nu} + (1-\eta)(\pi_t^o)^{1-\nu} \right]^{1/(1-\nu)}. \tag{6}$$

2.2 The Representative Household

The economy is populated by a representative household that derives utility from consumption

$$C_t = \left[\int_0^1 C_t(j)^{1-\nu} dj \right]^{1/(1-\nu)}$$

and real balances M_t/P_t, and disutility from working:

$$U_t = E_t \left[\sum_{s=0}^{\infty} \beta^s \left(\frac{(C_{t+s}/A_{t+s})^{1-\frac{1}{\tau}} - 1}{1 - \frac{1}{\tau}} + \chi \log \frac{M_{t+s}}{P_{t+s}} - h_{t+s} \right) \right], \tag{7}$$

where τ is the intertemporal substitution elasticity, χ is a scale factor, and h_{t+s} is hours worked. To ensure a balanced growth path, consumption in the utility function is adjusted for the level of technology, which can be interpreted as an aggregate habit stock.

The household supplies perfectly elastic labor services to the firms period by period for which it receives the real wage W_t. The household has access to a domestic capital market where nominal government bonds B_t are traded that pay (gross) interest R_t. Furthermore, it receives aggregate residual profits Π_t from the firms and has to pay lump-sum taxes T_t. Consequently, the household maximizes (7) subject to its budget constraint:

$$C_t + \frac{B_t}{P_t} + \frac{M_t}{P_t} + \frac{T_t}{P_t} = W_t h_t + \frac{M_{t-1}}{P_t} + R_{t-1} \frac{B_{t-1}}{P_t} + \Pi_t. \tag{8}$$

The usual transversality condition on asset accumulation applies which rules out Ponzi-schemes.

2.3 Fiscal Policy

The fiscal authority follows a 'passive' rule by endogenously adjusting the primary surplus to changes in the government's outstanding liabilities. It is assumed that the government levies a lump-sum tax (or subsidy) T_t/P_t to finance any shortfall in government revenues (or to rebate any surplus):

$$\frac{T_t}{P_t} = G_t + \frac{M_t - M_{t-1}}{P_t} - \frac{B_t - R_{t-1} B_{t-1}}{P_t}. \tag{9}$$

G_t denotes aggregate government purchases. We assume for simplicity that the government consumes a fraction of each individual good: $G_t(j) = \zeta_t Y_t(j)$. Define $g_t = 1/(1-\zeta_t)$. The aggregate demand can then be written in log-linear form as

$$\ln Y_t = \ln C_t + \ln g_t, \tag{10}$$

where $\ln g_t$ is assumed to be an exogenous process of the form

$$\ln g_t = (1-\rho_g)\ln g + \rho_g \ln g_{t-1} + \epsilon_{g,t}. \tag{11}$$

The process $\ln g_t$ can more generally be interpreted as a shift that affects the relationship between output and the marginal utility of consumption.

2.4 Monetary Policy

The central bank supplies money to the economy to control the nominal interest rate. We assume that it systematically reacts to the inflation rate according to the rule

$$R_t = (R_t^*)^{1-\rho_R} R_{t-1}^{\rho_R} \exp\{\epsilon_{R,t}^*\}, \tag{12}$$

where the desired nominal interest rate is given by

$$R_t^* = (r\pi_t^*) \left(\frac{\pi_t}{\pi_t^*}\right)^{\psi}. \tag{13}$$

Here r denotes the steady-state real rate, ρ determines the degree of interest rate smoothing, and ψ is the elasticity of the desired interest rate with respect to the deviation of inflation from its target π_t^*. The monetary policy shock $\epsilon_{R,t}^*$ captures unanticipated deviations from the systematic component of the policy rule due to a deliberate choice by the policy maker or an implementation error.

While in most of the literature the target rate π_t^* is assumed to be constant over time and known to the public, we assume in this paper that it is stochastic from the perspective of the public. Our model provides no explanation for the central bank's choice of π_t^*. We assume that there are periods in time when the desired inflation is high, $\pi_t^* = \pi_H^*$, and periods when it is low, $\pi_t^* = \pi_L^*$. More specifically, the log target rate evolves according to a two-state Markov-switching process:

$$\ln \pi_t^*(s_t) = \begin{cases} \ln \pi_L^* & \text{if} \quad s_t = 1 \\ \ln \pi_H^* & \text{if} \quad s_t = 2 \end{cases} \tag{14}$$

with transition matrix

$$\mathcal{P} = \left[\begin{array}{cc} \phi_1 & 1-\phi_2 \\ 1-\phi_1 & \phi_2 \end{array} \right],$$

where ϕ_j is the transition probability $I\!P[s_t = j | s_{t-1} = j]$. The expected value of the log target inflation rate is

$$\ln \pi^* = \frac{1-\phi_2}{2-\phi_1-\phi_2} \ln \pi_L^* + \frac{1-\phi_1}{2-\phi_1-\phi_2} \ln \pi_H^*. \tag{15}$$

Since it has been emphasized by some authors, e.g., Sims (2000a), that the variance of the policy shock $\epsilon_{R,t}^*$ has changed over time, we let the standard deviation $\sigma_R(s_t)$ be a function of the regime. The state s_t is unobserved by the public and hence agents have to learn from past inflation and interest rates whether they are in a high inflation or low inflation regime.

2.5 Detrending, Steady States, and Log-linear Approximation

To induce stationarity we detrend consumption, output, wages, and real money balances by the level of technology, A_t. In terms of the detrended variables, the model economy has a unique steady state. The system is in steady-state if the shocks $\epsilon_{z,t}$, $\epsilon_{g,t}$, and $\epsilon_{R,t}^*$ are zero and there is no uncertainty about the inflation target, that is, $\pi_L^* = \pi_H^* = \pi^*$. It is straightforward to verify from the first-order conditions of the optimization problems (4) and (7), not reported here, that the steady-state relationships are given by

$$\pi = \pi^*, \quad r = \frac{e^\gamma}{\beta}, \quad w = 1 - \frac{1}{\nu}, \quad c = \left(1 - \frac{1}{\nu}\right)^{1/\tau}, \quad y = g\left(1 - \frac{1}{\nu}\right)^{1/\nu}.$$

There is no trade-off between output and inflation in the steady state.

Define the deviations of output, inflation, nominal interest rate, and government spending from their respective steady states as

$$\widetilde{y}_t = \ln \frac{Y_t}{yA_t}, \quad \widetilde{\pi}_t = \ln \frac{\pi_t}{\pi}, \quad \widetilde{R}_t = \ln \frac{R_t}{R}, \quad \widetilde{g}_t = \ln \frac{g_t}{g}.$$

A log-linear approximation of the market-clearing conditions and the first-order conditions of the household's and firms' problems leads to the following two equations:

$$\widetilde{y}_t = I\!E_t[\widetilde{y}_{t+1}] - \tau(\widetilde{R}_t - I\!E_t[\widetilde{\pi}_{t+1}]) - I\!E_t[\Delta\widetilde{g}_{t+1}] + \tau I\!E_t[z_{t+1}], \tag{16}$$

$$\widetilde{\pi}_t = \frac{e^{\gamma}}{r} I\!E_t[\widetilde{\pi}_{t+1}] + \kappa[\widetilde{y}_t - \widetilde{g}_t]. \tag{17}$$

Equation (16) is an intertemporal consumption Euler equation. The expectational Phillips curve (17) is derived from the firms' optimal price-setting decision and governs inflation dynamics around the steady state. The parameter κ is a function of the probability that firms can re-optimize their price. In the absence of stickiness $kappa = 0$. The term $[\widetilde{y} - \widetilde{g}_t]/\tau$ corresponds to the marginal costs of production. Since $\widetilde{g}_t$ in this model links output to the marginal utility of consumption it can be broadly interpreted as preference shifter.

While (16) and (17) are standard components of New Keynesian models, we now turn to the approximation of the policy rule. Our derivation follows the approach in Andolfatto, Moran, and Hendry (2002). Divide (12) by the steady-state nominal rate $R = r\pi$ and take logs

$$\ln \frac{R_t}{R} = (1-\rho_R) \ln \frac{\pi^*_t}{\pi} + (1-\rho_R)\psi[\ln \frac{\pi_t}{\pi} - \ln \frac{\pi^*_t}{\pi}] + \rho_R \ln \frac{R_{t-1}}{R} + \epsilon^*_{R,t}.$$

Thus,

$$\widetilde{R}_t = (1-\rho_R)\psi\widetilde{\pi}_t + \rho_R\widetilde{R}_{t-1} + \epsilon_{R,t}, \tag{18}$$

where the composite monetary policy shock $\epsilon_{R,t}$ is defined as

$$\epsilon_{R,t} = (1-\rho_R)(1-\psi)\widetilde{\pi}^*_t + \epsilon^*_{R,t}$$

and $\widetilde{\pi}^*_t = \ln(\pi^*_t/\pi)$.[2] The agents face a signal extraction problem, as they observe the composite shock $\epsilon_{R,t}$ at time t, but are uncertain about the target rate $\widetilde{\pi}^*_t$. They

[2]Equation (18) is less general than the specification in Sims (2000a) who also allows for time-variation in ρ_R and ψ.

will use a Bayesian learning rule to update their beliefs about $\widetilde{\pi}_t^*$. Equations (3), (11), (14), (16) - (18), together with a learning rule for $\widetilde{\pi}_t^*$ determine the evolution of output, inflation, and nominal interest rates.

Recent work on the stability of monetary policy reaction functions, e.g., Clarida, Galí, and Gertler (2000), has suggested that the elasticity of the interest rate target with respect to inflation deviations has changed at the beginning of the 1980's during Paul Volcker's tenure as chairman of the Board of Governors. However, in a log-linear approximation of R_t^*, the effect of a change in ψ is of smaller order than the effect of a shift in the target inflation rate π_t^* and therefore not explored in this paper. Lubik and Schorfheide (2002) estimate a New Keynesian DSGE model based on subsamples and allow for a simultaneous change in the target inflation rate as well as the elasticity ψ but without taking 'learning' effects into account.

2.6 Model Solution

We have to solve the linear rational expectations system that consists of Equations (16) to (18), the law of motion for technology, government spending, and the Markov process for the inflation target. Define the vector of relevant model variables $x_t = [\widetilde{y}_t, \widetilde{\pi}_t, \widetilde{R}_t, I\!\!E_t[\widetilde{y}_{t+1}], I\!\!E_t[\widetilde{\pi}_t], y_{t-1}, \widetilde{g}_t, \widetilde{z}_t]'$, the vector of exogenous shocks $\epsilon_t(s_t) = [\epsilon_{g,t}, \epsilon_{z,t}, \epsilon_{R,t}(s_t)]'$, and the vector of expectation errors $\eta_t = [(\widetilde{y}_t - I\!\!E_{t-1}[\widetilde{y}_t]), (\widetilde{\pi}_t - I\!\!E_{t-1}[\widetilde{\pi}_t])]'$, where the expectations are taken with respect to the information set of the agents. The log-linearized model can be written as

$$\Gamma_0 x_t = \Gamma_1 x_{t-1} + \Psi\epsilon_t + \Pi\eta_t \tag{19}$$

and has a solution of the form (see Sims (2000b)):

$$x_t = \Theta_1 x_{t-1} + \Theta_0\epsilon_t + \Theta_x \sum_{j=1}^{\infty} \Theta_f^{j-1}\Theta_\epsilon I\!\!E_t[\epsilon_{t+j}]. \tag{20}$$

While the shocks to the technology and government spending process are assumed to be martingale differences, the composite technology shock $\epsilon_{R,t}$ has non-zero conditional expectations.

Based on the model specification all information that agents have about the state of monetary policy, s_t, is contained in the sequence of composite monetary policy shocks $\epsilon_R^t = \{\epsilon_{R,1}, \ldots, \epsilon_{R,t}\}$. Let $\xi_t = [1,0]'$ if $s_t = 1$ and $\xi_t = [0,1]'$ otherwise. Denote the estimate of ξ_t based on time t information by

$$\hat{\xi}_{t|t} = \begin{bmatrix} I\!P[s_t = 1|\epsilon_R^t] \\ I\!P[s_t = 2|\epsilon_R^t] \end{bmatrix}. \tag{21}$$

Future monetary policy regimes, ξ_{t+j}, can be forecasted based on the Markov transition matrix $\mathcal{P}$:

$$\hat{\xi}_{t+j|t} = \mathcal{P}^j \hat{\xi}_{t|t}. \tag{22}$$

Under the assumption that the monetary policy shock $\epsilon_{R,t}^*$ is normally distributed with regime-dependent standard deviation $\sigma_R(s_t)$ the regime conditional distribution of $\epsilon_{R,t+1}$ is

$$\epsilon_{R,t+1}|s_t \sim \mathcal{N}\Big((1-\rho_R)(1-\psi)\widetilde{\pi}(s_t), \sigma_R^2(s_t)\Big). \tag{23}$$

Once the new composite policy shock $\epsilon_{R,t+1}$ has been observed, beliefs about the state of monetary policy can be updated using Bayes theorem. Let ζ_t be the 2×1 vector that stacks the conditional densities $p(\epsilon_{R,t+1}|s_t = j)$, $j = 1,2$. Then

$$\hat{\xi}_{t+1|t+1} = \frac{\hat{\xi}_{t+1|t} \odot \zeta_t}{\hat{\xi}'_{t+1|t}\zeta_t}, \tag{24}$$

where $\odot$ denotes element-by-element multiplication.

The conditional expectations of the composite shock are given by

$$I\!E_t[\epsilon_{R,t+j}] = (1-\rho_R)(1-\psi)[\widetilde{\pi}_L^*, \widetilde{\pi}_H^*]\mathcal{P}^j\hat{\xi}_{t|t}(\epsilon_R^t). \tag{25}$$

Overall, the solution that takes into account the agents' learning about the state of monetary policy has the structure

$$x_t = \Theta_1 x_{t-1} + \Theta_0 \epsilon_t(s_t) + f_l\big(\epsilon_R^t(s_t)\big). \tag{26}$$

The solution can be computed recursively. It depends on s_t only through the composite monetary policy shock $\epsilon_{R,t}$. However, the solution is a function of the entire history of shocks. The term $f_l(\epsilon_R^t)$ arises through the learning process of the agents.

In our empirical analysis we will contrast the 'learning' solution of the monetary business cycle model with a 'full-information' solution in which the current state of monetary policy, s_t, is known by households and firms. This knowledge could stem from a credible announcement of the current policy regime by the central bank. However, there is uncertainty about future states of monetary policy. In this case the conditional expectations of the composite monetary policy shock are given by

$$\mathbb{E}_t[\epsilon_{R,t+j}] = (1-\rho_R)(1-\psi)[\widetilde{\pi}^*_L, \widetilde{\pi}^*_H]\mathcal{P}^j \xi_t \tag{27}$$

and the law of motion of x_t is of the form

$$x_t = \Theta_1 x_{t-1} + \Theta_0 \epsilon_t(s_t) + f_f(s_t). \tag{28}$$

3 Econometric Approach

The model presented in the previous section is fitted to observations on output growth, inflation, and the nominal interest rate, stacked in the vector y_t. The observables y_t can be expressed as a linear function of the model variables x_t

$$y_t = A_0 + A_1 x_t. \tag{29}$$

Equations (14), (26) or (28), and (29) provide a state-space model for y_t with regime switching. We assume that the vector of shocks ϵ_t has a normal distribution with diagonal covariance matrix conditional on the regime s_t. The system matrix of this state space model depend on the vector of structural parameters

$$\theta = [\gamma, \pi^*_L, \pi^*_H, r^*, \tau, \kappa, \psi, \rho_g, \rho_z, \rho_R, \sigma_g, \sigma_z, \sigma_{R,L}, \sigma_{R,H}]',$$

where σ_g, σ_z, $\sigma_{R,L}$, $\sigma_{R,H}$ are the standard deviations of the structural shocks. The transition probabilities are stacked in the vector $\phi = [\phi_1, \phi_2]'$, and the histories of y_τ, x_τ, s_τ, $\tau = 1, \ldots, t$ are denoted by Y^t, X^t, and S^t, respectively. We will place a prior distribution on the parameter vectors θ and ϕ and conduct Bayesian inference, described in Section 3.1. Section 3.2 will focus on the computation of the likelihood function for the state space model and our strategy to generate draws from the posterior distribution of θ, ϕ and the history of latent states S^T. The use of posterior probabilities as a tool for model comparisons is discussed in Section 3.3.

3.1 Posterior Inference

The analysis in the remainder of this paper is is conditional on an initial observation y_0. The latent state vector S^T is treated in the same way as the parameter vectors θ and ϕ. Once a prior distribution has been specified, Bayesian inference is conceptually straightforward. All the information about the parameters is summarized in the posterior distribution, which is obtained through Bayes theorem

$$p(\theta, \phi, S^T|Y^T) = \frac{p(Y^T|\theta, \phi, S^T)p(S^T|\phi)p(\phi, \theta)}{p(Y^T)}, \tag{30}$$

where $p(Y^T|\theta, \phi, S^T)$ is the likelihood function, $p(S^T|\phi)$ is the prior for the latent states – given by the Markov process (14), and $p(\phi, \theta)$ is the prior for θ and ϕ. The term in the numerator is often referred to as marginal data density and defined as

$$p(Y^T) = \int p(Y^T|\theta, \phi, S^T)p(S^T|\phi)p(\phi, \theta)d(\theta, \phi, S^T). \tag{31}$$

The practical difficulty is to characterize the posterior distribution. For the models at hand, we will describe an algorithm that allows us to generate draws of θ, ϕ, and S^T from this posterior distribution.

In many Markov-switching models it is convenient to use a Gibbs-sampling approach and draw iteratively from the following conditional posterior distributions (see Kim and Nelson (1999) for details):

$$p(S^T|\theta, \phi, Y^T), \quad p(\phi|\theta, S^T, Y^T), \quad \text{and} \quad p(\theta|\phi, S^T, Y^T),$$

because all three conditionals can be characterized in terms of well-known distributions that can be easily simulated.

For the model presented in the previous section, however, it is difficult to generate draws from $p(\theta|\phi, S^T, Y^T)$ due to the nonlinear relationship between θ and the system matrices of the state space model. Hence, we factorize the joint posterior as

$$p(\theta, \phi, S^T|Y^T) = p(\theta, \phi|Y^T)p(S^T|\theta, \phi, Y^T) \tag{32}$$

and use the Metropolis-Hastings algorithm described in Schorfheide (2000) to generate draws from $p(\theta, \phi|Y^T)$. Conditional on the latent states we use Kim's (1994) smoothing algorithm to generate draws from the history S^T of latent states. The smoothing algorithm is exact for the 'learning' specification and provides an approximation in the case of the 'full-information' specification.[3]

3.2 The Likelihood Function

To generate draws from the posterior distribution of θ and ϕ with the Metropolis-Hastings algorithm it is necessary to evaluate the likelihood function $p(Y^T|\theta,\phi)$. In linear Gaussian state space models without regime switching the likelihood function can be easily computed with the Kalman Filter (see, for instance, Hamilton (1994)). The Kalman filter tracks the means and variances of the following conditional distributions recursively (for the sake of brevity θ and ϕ are omitted from the conditioning set):

(i) Initialization: $p(x_t|Y^t)$.

(ii) Forecasting: $p(x_{t+1}|Y^t) = \int p(x_{t+1}|x_t)p(x_t|Y^t)dx_t$,
$p(y_{t+1}|Y^t) = \int p(y_{t+1}|x_{t+1},Y^t)p(x_{t+1}|Y^t)dx_{t+1}$.

(iii) Updating: $p(x_{t+1}|Y^{t+1}) = p(y_{t+1}|x_{t+1},Y^t)p(x_{t+1}|Y^t)/p(y_{t+1}|Y^t)$.

The likelihood function is obtained from

$$p(Y^T|\theta,\phi) = \prod_{t=1}^{T} p(y_t|Y^{t-1},\theta,\phi). \tag{33}$$

Unfortunately, the presence of a latent Markov state s_t complicates the computation of the various conditional distributions. We will distinguish between the 'learning' and the 'full-information' specification.

3.2.1 Learning Specification

The policy rule has a special structure that allows us to recover the composite monetary policy shocks directly from the observables conditional on θ and ϕ:

$$\begin{aligned} \epsilon_{R,t}(s_t) &= \ln R_t - \rho_R \ln R_{t-1} - (1-\rho_R)\psi \ln \pi_t \\ &\quad -(1-\rho_R)\ln r - (1-\rho_R)(1-\psi)\ln \pi^*. \end{aligned} \tag{34}$$

[3] The accuracy of this approximation does not affect our inference with respect to θ and ϕ.

Since x_t in Equation (26) depends on s_t only through $\epsilon_{R,t}$, we can deduce that x_t provides no information about s_t conditional on the observables Y^t. Thus, at the initialization step of the filter

$$p(s_t, x_t|Y^t) = p(x_t|Y^t)p(s_t|Y^t). \tag{35}$$

Define the vector $\lambda(\theta)$ such that

$$\begin{aligned} \lambda' y_t &= \ln R_t - \rho_R \ln R_{t-1} - (1-\rho_R)\psi \ln \pi_t \\ &= (1-\rho_R)\ln r + (1-\rho_R)(1-\psi)\ln \pi^* + \epsilon_{R,t}. \end{aligned} \tag{36}$$

We factorize the distribution of y_{t+1} given Y^t as

$$p(y_{t+1}|Y^t) = p(\lambda' y_{t+1}|Y^t)p(y_{t+1}|\lambda' y_{t+1}, Y^t) \tag{37}$$

and examine the two components separately. According to Equation (36) the history of X^t provides no additional information about $\lambda' y_{t+1}$ conditional on Y^t. Thus, it is straightforward to compute

$$p(\lambda' y_{t+1}|Y^t) = \sum_{j=1}^{2} p(\lambda' y_{t+1}|s_{t+1} = j, Y^t) I\!P[s_{t+1} = j|Y^t], \tag{38}$$

where $I\!P[s_{t+1} = j|Y^t]$ can be obtained from Equation (22). The beliefs about the latent Markov state s_{t+1} can be updated according to Equation (24), which yields $p(s_{t+1}|Y^{t+1})$. Note that all the information in y_{t+1} with respect to s_{t+1} is contained through $\epsilon_{R,t+1}$ in the linear combination $\lambda' y_{t+1}$. To analyze the second term in (37) recall that s_{t+1} provides no information about x_{t+1} conditional on $\lambda' y_{t+1}$ and Y^t. Therefore, one can evaluate

$$p(y_{t+1}|\lambda' y_{t+1}, Y^t) = \int p(y_{t+1}|\lambda' y_{t+1}, x_{t+1}, Y^t)p(x_{t+1}|\lambda' y_{t+1}, Y^t)dx_{t+1} \tag{39}$$

and

$$p(x_{t+1}|Y^{t+1}) = p(y_{t+1}|\lambda' y_{t+1}, x_{t+1}, Y^t)p(x_{t+1}|\lambda' y_{t+1}, Y^t)/p(y_{t+1}|Y^t) \tag{40}$$

using the standard formulas of the Kalman filter.

3.2.2 Full-Information Specification

The evaluation of the likelihood function for the 'full-information' specification is more complicated since the Markov state s_t enters the conditional distribution of x_t directly (see Equation 28). In this case, the distribution of x_t given the trajectory Y^t is a mixture of 2^t components and it is even for very small sample sizes computationally cumbersome to keep track of all the mixture components. As noted in the literature (see Kim and Nelson (1999) for a survey), the key to keeping the filter operable is to collapse some of the mixture components at the end of each iteration. We are using an algorithm that approximates $p(x_t|Y^t)$ by a mixture with 2^k components.[4] For each component and conditional on $s_{t+1} = 1$ and $s_{t+1} = 2$ the standard Kalman filter iterations are used to compute

$$\begin{aligned}
p(x_{t+1}|Y^t) &= \sum_{j=1}^{2} p(x_{t+1}|Y^t, s_{t+1} = j) I\!P[s_{t+1} = j|Y^t] \\
p(y_{t+1}|Y^t) &= \sum_{j=1}^{2} p(y_{t+1}|Y^t, s_{t+1} = j) I\!P[s_{t+1} = j|Y^t] \\
I\!P[s_{t+1} = j|Y^{t+1}] &= p(y_{t+1}|Y^t, s_{t+1} = j) I\!P[s_{t+j} = j|Y^t]/p(y_{t+1}|Y^t) \\
p(x_{t+1}|Y^{t+1}) &= \sum_{j=1}^{2} p(x_{t+1}|Y^{t+1}, s_{t+1} = j) I\!P[s_{t+1} = j|Y^{t+1}].
\end{aligned}$$

The updated distribution $p(x_{t+1}|Y^{t+1})$ is a mixture with 2^{k+1} components, half of which we eliminate. Components with approximately zero weight are removed and pairs of components that are based on similar histories of $s_{t+1}, \ldots, s_{t-k+2}$ are aggregated into a single component with the same mean and variance.[5]

3.3 Posterior Model Probabilities

In the subsequent empirical analysis we are interested in comparing different specifications of the model economy. We do this by computing posterior odds. The

[4] The results in Section 4 are based on $k = 3$. To assess the accuracy of the fixed-mixture approximation we evaluated the likelihood function at the posterior mode for k ranging from 3 to 20. The log likelihood differentials were less than 0.01.

[5] Details of the algorithm are provided in a Technical Appendix that is available upon request.

posterior odds of a specification $\mathcal{M}_0$ versus a specification $\mathcal{M}_1$ are given by[6]

$$\frac{\pi_{0,T}}{\pi_{1,T}} = \left(\frac{\pi_{0,0}}{\pi_{1,0}}\right)\left(\frac{p(Y^T|\mathcal{M}_0)}{p(Y^T|\mathcal{M}_1)}\right). \tag{41}$$

The first factor is the prior odds ratio in favor of $\mathcal{M}_0$. The second term is called the Bayes factor and summarizes the sample evidence in favor of $\mathcal{M}_0$. The term $p(Y^T|\mathcal{M}_i)$ is called Bayesian data density and given by Equation (31). The logarithm of the marginal data density can be interpreted as maximized log-likelihood function penalized for model dimensionality, see, for instance, Schwarz (1978). We use a numerical technique known as modified harmonic mean estimation, proposed by Geweke (1999), to approximate the data density.[7]

4 Empirical Analysis

The log-linearized monetary DSGE model described in Section 2 is fitted to quarterly U.S. data on output growth, inflation, and nominal interest rates from 1960:I to 1997:IV.[8] The first step in the empirical analysis is the specification of a prior distribution for the structural parameters. Table 1 provides information about the distributional forms, means, and 90% confidence intervals. The model parameters are assumed to be independent *a priori*. The target inflation rates π^*_L and π^*_H, the steady-state real interest rate r, and the standard deviations of the monetary policy shock $\sigma_{R,L}$ and $\sigma_{R,H}$ are annualized. Since the solution of the linear rational expectations model may be non-existent or exhibit multiple equilibria, we truncate

[6]According to Jeffreys (1961) the posterior odds may be interpreted as follows: $\pi_{0,T}/\pi_{1,T} > 1$ null hypothesis is supported; $1 > \pi_{0,T}/\pi_{1,T} > 10^{-1/2}$ evidence against $\mathcal{M}_0$ but not worth more than a bare mention; $10^{-1/2} > \pi_{0,T}/\pi_{1,T} > 10^{-1}$ substantial evidence against $\mathcal{M}_0$; $10^{-1} > \pi_{0,T}/\pi_{1,T} > 10^{-3/2}$ strong evidence against $\mathcal{M}_0$; $10^{-3/2} > \pi_{0,T}/\pi_{1,T} > 10^{-2}$ very strong evidence against $\mathcal{M}_0$; $10^{-2} > \pi_{0,T}/\pi_{1,T}$ decisive evidence against $\mathcal{M}_0$.

[7]We actually found that Geweke's (1999) approach is more stable than the procedure proposed by Chib and Jeliazkov (2001).

[8]The time series are extracted from the DRI·WEFA database. Output growth is log difference of real per capita GDP (GDPQ), multiplied by 100 to convert into percent. Inflation is annualized percentage change of CPI-U (PUNEW). Nominal interest rate is average Federal Funds Rate (FYFF) in percent.

the joint distribution at the boundary of the indeterminacy region.[9] Our initial prior assigns about 15% probability to indeterminacy. Mean and confidence intervals are reported for the truncated version of the prior.

The prior confidence intervals for the parameters of the policy rule are fairly wide. The elasticity of the nominal interest rate with respect to inflation deviations from the the target lies between 1.0 and 2.2, whereas the interval for the smoothing coefficient ρ_R ranges from 0.2 to 0.8. We identify regime $s_t = 1$ as the low inflation regime with target rate π_L^* and impose that the differential $\pi_H^* \geq \pi_L^*$. The intervals for both $\ln \pi_L^*$ and $\ln \pi_H^* - \ln \pi_L^*$ range from 1.3 to 4.5 percent. The standard deviation of the monetary policy shock is, in both regimes *a priori* between 25 and 100 basis points. Our prior for the transition probabilities ϕ_1 and ϕ_2 implies that the duration of the policy regimes lies between 6 and 50 quarters. The prior for the annualized real interest rate is centered at 2 percent, which translates into a quarterly discount factor β of 0.995. The slope coefficient in the Phillips curve, κ, is chosen to be consistent with the range of values typically found in the New Keynesian literature (see, for instance, Galí and Gertler (1999)). The prior mean for the intertemporal substitution elasticity is $\tau = 0.5$.

4.1 Estimation of Parameters and Policy Regimes

We begin by estimating three specifications of the monetary DSGE model: the 'learning' specification given in (26), the 'full-information' specification (28), and a version of the model without regime switching in the monetary policy rule for which $\pi_H^* = \pi_L^*$ and $\sigma_{R,L} = \sigma_{R,H}$. Posterior means and confidence intervals are reported in Table 2. We will focus our discussion on the estimated policy parameters.

Most striking is the estimated difference in the target inflation rates: 4.8 percent in the BL specification and 5.2 percent under full information. The estimated probabilities of a regime shift are small, 5 percent or less, which means that the duration

[9] Lubik and Schorfheide (2002) estimate a version of the model without Markov switching based on subsamples and allow for the possibility of indeterminacy and sunspot driven business cycle fluctuations.

of the regimes is more than 5 years. The Bayes factor reported in Table 3 clearly favors the BL specification over the no-regime-shift version of the DSGE model.

To gain better insight in the evolution of monetary policy over time, we plot the smoothed posterior probabilities of the high-inflation-target regime in Figure 1 together with the NBER business cycle dates. The graph is consistent with the view that policy makers in the late 1960's and early 1970's attempted to exploit a Phillips curve relationship by raising the inflation target in order to achieve lower unemployment. The high inflation episode ended with Volcker's stabilization in the early 1980's after growing scepticism about an exploitable long-run trade-off between unemployment and inflation.[10] Of course, our analysis itself cannot rule out other explanations of the shift in the target rate. The 'learning' estimates suggests that the Fed shifted to the high-inflation regime during the 1970 recession , shortly lowered the target between 1971 and 1973 and raised it again subsequently.

Our findings are in sharp contrast to the results reported in Sims (2000a) and Sims and Zha (2002). Both papers examine the evolution of monetary policy with autoregressive Markov-switching models. Sims (2000a) estimates a univariate reaction function with regime-dependent coefficients and finds no evidence in favor of two-time shifts in the early 1970's and the beginning of the 1980's. Instead, his estimates, as well as the multivariate analysis in Sims and Zha (2002), suggest frequent switching between regimes throughout the post-war period.

To understand the source of the discrepancy it is useful to consider the composite monetary policy shock

$$\epsilon_{R,t} = (1-\rho)(1-\psi)\tilde{\pi}^*_t + \epsilon^*_{R,t}.$$

According to the posterior mean estimates of the 'learning' specification, the term of $\epsilon_{R,t}$ that is due to the shifting inflation target takes the value 0.33 with probability 0.63 and -0.55 with probability 0.37. Thus, its standard deviation is about 0.43, which is smaller than the estimated standard deviation of $\epsilon^*_{R,t}$. Based on a univariate

[10]Sargent (1999) offers two explanations for this scepticism: the 'triumph of the natural rate theory' and the 'vindication of econometric policy evaluation'.

estimation of the policy function it is very difficult to detect shifts in $\tilde{\pi}_t^*$.[11] Hence, the identification of the regimes comes effectively from the inflation equation of the DSGE model, not from the policy rule. Since the model does not imply time-variation in the price-setting behavior of the firms, nor allows for fundamental shifts in the marginal cost process (except through output variation and the $\widetilde{g}_t$ process), the observed rise in inflation and the subsequent decline are attributed to two policy regime shifts.

The estimates in Table 2 indicate that the high-inflation-target regime coincides with a high volatility of the monetary policy shock and more erratic behavior of the Fed.[12] To assess the importance of time-variation in the size of the disturbance to the policy rule we re-estimate the 'learning' and 'full-information' specifications under the homoskedasticity restriction $\sigma_{R,L} = \sigma_{R,H}$. According to the posterior odds reported in Table 3 the homoskedasticity restriction is rejected both under 'learning' and 'full information'. In fact, the 'learning' specification with homoskedastic policy shocks is dominated by the version of the DSGE model that is not subject to regime shifts.

A question that has received a lot of attention in the recent literature is whether post-war monetary policy has been active or passive. A monetary policy is considered active if it involves strong enough eventual reaction of the interest rate to the inflation rate to guarantee a unique equilibrium. While the estimation method used in this paper rules out regions of the parameter space that lead to equilibrium indeterminacy it is nevertheless interesting to examine the estimates of ψ. Both in the 'learning' and 'full-information' specification $\hat{\psi}$ is about 1.7 and the confidence interval is clearly bounded away from one. In the absence of regime switching, the estimate of ψ is much closer to the indeterminacy region as the confidence interval ranges from 1 to 1.3. Inflation enters the policy reaction function through the term $\psi \ln(\pi_t/\pi_t^*)$. The interpretation of the experience in the 1970's provided by

[11] In fact, a univariate estimate of the reaction function used in this paper leads to frequent rather than two-time shifts.

[12] To some extent the policy shock might capture changes in the systematic part of monetary policy that are omitted from our model.

the no-regime-shift model is that the deviation from the inflation target was large while ψ was small. The regime switching estimates on the other hand, indicate that the target inflation rate was high in the 1970's and the observed interest rate movements are not inconsistent with a ψ that is substantially larger than one.[13] The regime-switching model studied here ignores possible time variation in ψ. However, the analysis in Lubik and Schorfheide (2002) does suggest that even after adjusting for different target inflation rates through sample splitting there is evidence in favor of passive monetary policy and equilibrium indeterminacy in the 1970's.

4.2 Evidence on the Importance of Learning

While in most monetary DSGE models it is assumed that the agents have full information about the state of monetary policy, the goal of this paper is to study the effect of learning about policy regimes on the dynamics of output growth, inflation, and interest rates. Since we have fitted both the 'learning' and 'full-information' specification to the data we can compare the two using the Bayes factor reported in Table 3. Under the assumption that the two specifications have equal prior probability, the Bayes factor implies that the posterior odds are 314 to 1 in favor of the 'full-information' specification. Thus, overall, incorporating the learning mechanism into the model does not lead to an improvement of fit. However, Bayesian marginal data densities and Bayes factors provide only a broad measure of model fit that captures one-step-ahead predictive performance.[14] Hence, it is useful to take a closer look at the dynamics generated by the learning mechanism.

We will study Volcker's disinflation policy through the lens of the monetary DSGE model. In addition to the 'learning' and 'full-information' specifications we also estimate a 'no-learning' specification of the form

$$X_t = \Theta_1 x_{t-1} + \Theta_0 \epsilon_t(s_t) \tag{42}$$

[13] This point was raised by Eric Leeper in response to the results reported in Lubik and Schorfheide (2002).

[14] Since $\ln p(Y^T|\mathcal{M}) = \sum_{t=1}^{T} \ln p(y_t|Y^{t-1}, \mathcal{M})$ the data density can be interpreted as predictive score.

in which the agents ignore the serial correlation in the composite shock $\epsilon_{R,t}$ when they form their expectations. The disinflation corresponds to a shift from the high-inflation regime, $s_{T_*} = 2$, to the low-inflation regime, $s_{T_*+h} = 1$, $h = 1, \ldots$. According to our regime estimates the transition occurs between 1982:III and 1982:IV. However, these estimates seem to capture the end rather than the beginning of the disinflation period. Hence, we choose T_* to be 1980:I in the subsequent experiment. For each model specification $\mathcal{M}_i$ we use the appropriate filter described in Section 3 to compute $p(x_{T_*}|Y^{T_*})$. For the BL model we also have to compute $I\!P[s_{T_*} = 2|Y^{T_*}]$ which is close to one by the beginning of 1980. We then set $s_{T_*+1} = \ldots = s_{T_*+12} = 1$ (low inflation target) and compute expected values for $x_{T_*+1}, \ldots, x_{T_*+12}$.[15] The results are depicted in Figure 2.

Under the assumption that the disinflation is announced by the Central Bank and that the announcement is credible, the probability of the high-inflation regime drops from one to zero. If agents are Bayesian learners, skeptical about policy announcements, the belief about π_t^* change gradually. After two years the probability of the high-inflation regime has dropped to about 10 percent. In the 'no-learning' scenario agents' beliefs correspond to the estimated steady state regime probabilities. The beliefs about the policy regimes are reflected in the predicted path of expected inflation. Under 'full information', expected inflation drops rapidly from 12 to 5 percent, whereas under the assumption of 'learning' the decay is sluggish and expected inflation stays above 7 percent until 1983.

According to the price setting equation current inflation is an increasing function of expected inflation. Hence, the drop in the latter is associated with a fall of the inflation rate. Under 'full information' the nominal rate falls immediately, whereas it rises initially under 'learning'. The decay of the nominal interest rate is delayed through the interest rate smoothing. Overall, the transition to low inflation and nominal interest rate is much quicker under full information. This result is

[15] Since the learning model is non-linear in $\epsilon_{R,t}(s_t)$ we generate random draws of $\epsilon_{R,t}$ and average the trajectories of x_t. For all three specifications we average over θ and ϕ with respect to the posterior distributions $p(\theta, \phi|Y^T)$.

consistent with the findings reported by Andolfatto and Gomme (2003). During the disinflation period the inflation forecast bias is larger under ‘learning’ than under ‘full information’, a point that has been emphasized by Andolfatto, Moran, and Henry (2002). However, the bias is with less than 1 percent, fairly small in our analysis. The ‘learning’ version of the DSGE model predicts a rise of the ex-post real interest from 4 percent in 1980:I to about 9 percent in 1980:III and a subsequent decay. Under ‘full-information’ the real rate drops almost monotonically, starting from 9 percent in 1980:I.

Output growth is slightly negative in the first quarter of 1980 before it rises to 1 percent in the second quarter. Under ‘learning’ output growth starts to rise immediately. It peaks in 1981 and falls to about 0.5 percent by 1983. Conditional on information up to 1980:I our model predicts positive technology growth rates for the subsequent quarters, which offset the contractionary effect of the disinflation policy.

In the data the decline of inflation and the nominal interest rate was indeed sluggish. Figure 2 suggests that the updating of beliefs about current and future monetary policy may have delayed the disinflation. With the exception of a drop in the third quarter of 1980, the ex-post real rate rises from about 2 percent in 1980:I to more than 10 percent in 1982:III and slowly falls afterwards. Thus, its path vaguely resembles the hump-shaped response of the ‘learning’ specification. However, none of the DSGE model specifications is able to predict the third quarter drop of the interest rates. While the full-information version of the DSGE model predicts a slight recession at the beginning of 1980, the actual drop in output growth was much larger. Moreover, the assumed average output growth over the 12 periods is smaller than the growth predicted by any of the specifications.

4.3 Learning and Policy Interventions

The effects of monetary policy in dynamic models are often predicted with impulse responses to monetary policy shocks, i.e., unanticipated deviations from the policy reaction function. However, if a policy intervention is sustained over a long period in time, agents are likely to interpret the policy as a regime shift to a 'new' policy rule rather than a one-time deviation from the 'old' policy rule and the intervention generates an expectation-formation effect of the kind that Lucas (1976) emphasized. The regime-switching model estimated in this paper allows us to quantitatively assess the importance of this expectation-formation effect.

The subsequent analysis is closely related to work by Leeper and Zha (2002). Leeper and Zha consider a calibrated two-equation model of output and money, in which money evolves according to a no-feedback money growth rate that is subject to Markov-switching. They model interventions through a sequence of policy shocks and define the expectation-formation effect of a policy intervention as the difference between the total effect and the effect under the assumption that agents know the actual regime. In our setup the former corresponds to the responses in the 'learning' specification, whereas the latter is equivalent to the 'full-information' response. Leeper and Zha argue based on their model that the expectation-formation effect is likely to be small whenever the effect of the intervention lies within two standard deviations of the policy effects that have been observed historically in a particular regime. Such interventions are called modest. The authors use an identified VAR to assess whether the interventions of the Federal Reserve Bank in the past have been modest.

We will use our estimated DSGE model to calculate the magnitude of the expectation-formation effect directly. We consider two types of interventions. The first intervention consists of a one-period interest increase of 25 basis points, the second intervention raises the interest rate by 100 basis points for five consecutive periods. To compute the responses we assume that the system is initially in the steady state and then construct the sequence of policy shocks that leads to the desired interest path. Formally, the responses are given by[16]

$$\mathbb{E}[y_t|x_0, \{y_{3,\tau}\}_{\tau=1}^{h}], \quad t = 1, \ldots, 20.$$

[16] As before, we average over θ and ϕ with respect to the appropriate posterior distributions.

Thus, we also integrate over s_t, with the following exception: to compute the responses for the full information specification in the second policy scenario, we assume that the policy is accompanied by a period-by-period announcement of a low inflation target.

Figure 3 depicts the responses to the interventions for 'learning', 'full-information', and 'no-learning'. In the first experiment, the full-information and the 'no-learning' solutions differ by the term $\Theta_f(s_t)$ which has expected value zero. Hence, both responses are identical. The most striking difference between the 'learning' and the 'full-information' responses is the magnitude of the intervention effect on output growth and inflation. Consider the log-linearized reaction function:

$$\widetilde{R}_t = (1-\rho_R)\psi\widetilde{\pi}_t + \rho_R\widetilde{R}_{t-1} + \epsilon_{R,t}$$

A positive policy shock $\epsilon_{R,t}$ lowers the demand for money has the tendency to reduce inflation. Hence, the shock $\epsilon_{R,t}$ has to exceed the desired nominal interest rate. In the learning model a positive $\epsilon_{R,t}$ leads agents to believe that the probability of the low-inflation-target regime has increased (to more than 75 percent). Thus, expected inflation drops and current falls further. The shock $\epsilon_{R,t}$ associated with a one-percent increase in interest rates is much larger under 'learning'. This can be seen in the fourth panel of Figure 3. The magnitude of the composite shock $\epsilon_{R,t}$ can be compared to the estimated standard deviations $\sigma_{R,L}$ and $\sigma_{R,H}$ given in Table 2.

The second column of Figure 3 depicts the effects of the second intervention. In the 'full-information' case, it is assumed that the agents associate the intervention with a shift to the low-inflation-target regime. The 'learning' responses of output and inflation lie between the 'full-information' and the 'no-learning' response. The disinflation effect of the interest rate policy is strongest under 'full-information' and weakest under 'no-learning'. The two experiments suggest that ignoring the effect of 'learning' if it is indeed present can generate quantitatively misleading predictions about the effect of a policy even if the intervention itself appears to be small and short-lived.

A major difference between the model used by Leeper and Zha (2002) and the

New Keynesian model in this paper is that current inflation feeds back into the policy rule and the ‘learning’ effect $\Theta_f(\epsilon_R^t$ in Equation (26) is important for the determination of the composite policy shock $\epsilon_{R,t}$. In other words, the presence of learning effects the identification of the policy shock. Some of the differences between the responses disappear if a more accurate identification is used for the ‘no-learning’ specification. We consider the following approximating model:

$$x_t = \Theta_1 x_{t-1} + \Theta_0 \epsilon_t + \Theta_f(\epsilon_{R,t}). \tag{43}$$

This model has the property that it correctly predicts the initial effect $\partial x_t / \partial \epsilon_{R,t}$ of the policy shock but ignores the dependence of the ‘learning’ effect on the entire past history of shocks. Hence, the approximating model model provides a more accurate identification of the policy shock $\epsilon_{R,t}$ given x_t and x_{t-1} than the ‘no-learning’ specification. The responses for this approximating model are summarized in Figure 4. The discrepancy between the output and inflation responses is now much smaller and the adjustment of the identification assumption to the presence of ’learning’ leads to more accurate predictions of the BL response with the approximating model.

Due to the heteroskedasticity of the monetary policy shock and the Bayesian learning mechanism agents associate large shocks with the high-inflation regime and, vice-versa, small interventions with the low-inflation regime. In Figure 5 we compare the ‘learning’ responses with responses that are calculated by setting $\sigma_{R,H} = \sigma_{R,L}$. Consider the first experiment. Under heteroskedasticity the 25 basis point intervention is interpreted as evidence of the low inflation regime. Hence, upon impact of the shock, the probability associated with the high-inflation regime drops to about 20 percent. Under homoskedasticity this probability stays about 33 percent. Hence, the predicted drops in inflation and output growth is less severe and closer to the ‘full-information’ responses. In the second experiment the intervention is fairly large and hence interpreted as evidence in favor of the high-inflation regime. Consequently, the probability associated with high inflation is larger under heteroskedasticity than under homoskedasticity.

4.4 Robustness and Caveats

The posterior estimates of the persistence, ρ_g, of the $\widetilde{g}_t$ process reported in Table 2 is close to unity which suggests that the price-setting Equation (17) cannot endogenously generate the persistence in the inflation series. Galí and Gertler (1999) propose a hybrid Phillips curve that involves lagged inflation on the right-hand-side. In the context of the model presented in Section 2, such an equation can be derived from the assumption that a fraction ω of the firms, that we are unable to re-optimize their price, adjust their price charged in the previous period, $P_{t-1}(j)$, by the lagged inflation rate π_{t-1} rather than the steady state inflation rate π. The resulting Phillips curve is of the form

$$\widetilde{\pi}_t = \frac{\omega}{1+\omega e^{\gamma}/r}\pi_{t-1} + \frac{e^{\gamma}/r}{1+\omega e^{\gamma}/r} I\!E_t[\widetilde{\pi}_{t+1}] + \frac{\kappa}{1+\omega e^{\gamma}/r}[\widetilde{y}_t - \widetilde{g}_t]. \tag{44}$$

It involves an additional parameter ω and nests Equation (17) as the special case $\omega = 0$.

We place a diffuse prior on the share ω, that is centered at $\omega = 0.5$ and estimate the 'learning' and the 'full-information' version of the monetary DSGE model. The posterior estimates of ω are around 0.2. A comparison of the two specifications based on the Bayes factor, see Table 3, indicates that the 'full-information' version is preferred. However, overall the introduction of backward looking price setting does not lead to an improvement over the $\omega = 0$ versions of the DSGE model.

Following Erceg and Levin (2001), we also consider a model in which the central bank reacts to output growth in addition to inflation:

$$R_t^* = (r\pi_t^*)\left(\frac{\pi_t}{\pi_t^*}\right)^{\psi_1}\left(\frac{y_t/y_{t-1}}{e^{\gamma}}\right)^{\psi_2}. \tag{45}$$

Under 'learning' the posterior mean estimate of ψ_2 in terms of annualized output growth is 0.14, and under 'full information' the estimate increases to 0.48. According to the Bayes factor, the inclusion of output growth in the policy rule does not improve the fit of the 'learning' version. However, it does improve the fit of the 'full-information' version of the DSGE model. As before, the latter strictly dominates the 'learning' specification in terms of time series fit.

The monetary DSGE models that have been estimated in this paper are very stylized. Labor supply is inelastic and investment and capital accumulation, which provide an important intertemporal link, are not modelled. Nevertheless, it has become one of the benchmark monetary DSGE models in recent years. To provide a reference for the overall fit of the model we also estimate a constant-coefficient VAR(1) with Minnesota prior.[17] The data density, reported in Table 3 of the VAR is orders of magnitude larger than the 'learning' model, which documents the stylized nature of the structural model. Nevertheless, we believe that some interesting lessons have been learned from the empirical analysis presented in this section.

5 Conclusion

We have estimated a simple New Keynesian monetary DSGE model that has become a popular benchmark model for the analysis of monetary policy. Unlike in earlier econometric work, monetary policy follows a rule that is subject to regime shifts. While our model provides no explanation why these regime shifts occur, we assume that the public has potentially incomplete information about the state of monetary policy and has to learn about the current regime. Our regime estimates are consistent with the popular story that monetary policy is characterized by a high-inflation regime in the 1970's which ended with Volcker's stabilization policy in the early 1980's.

The evidence on the importance of learning and uncertainty about the policy regime is mixed. Posterior probabilities of the 'full information' versus 'learning' specification of the DSGE model favor the former. On the other hand, a closer look at the disinflation episode in the early 1980's indicates that the fall of inflation and interest rates is better explained by the delayed response of the 'learning' specification.

The presence of a learning mechanism has potentially important consequences for the prediction of the effects of policy interventions. A prolonged intervention

[17]The precise implementation is described in Del Negro and Schorfheide (2002).

might be interpreted by the agents as a shift to a new policy regime and lead to changes in the agents' expectation formation. We document these expectation formation effects with our estimated DSGE models. Ignoring the effect of the expectation formation can result in misleading predictions even if the intervention appears to be small, in particular if the predictions are generated from over-identified DSGE models. The prediction errors can potentially be reduced if the response to the monetary policy shock in the initial period is corrected to capture the contemporaneous effect of the monetary policy shock more accurately. Our analysis also raises the question whether agents associate large interventions with a shift back to a high inflation regime. If they do, then this can substantially alter the responses of output growth and inflation. Since many of the popular VAR identification schemes impose fairly weak restrictions among the contemporaneous variables they may well be consistent with the presence of a learning effect and deliver accurate predictions as argued by Leeper and Zha (2002).

References

Andolfatto, David, Kevin Moran, and Scott Hendry (2002): "Inflation Expectations and Learning about Monetary Policy," *Manuscript*, Bank of Canada.

Andolfatto, David and Paul Gomme (2003): "Monetary Policy Regimes and Beliefs," *International Economic Review*, **44**, 1-30.

Calvo, G.A. (1983): "Staggered Prices in a Utility Maximizing Framework," *Journal of Monetary Economics*, **12**, 382-398.

Chib, Siddhartha and Ivan Jeliazkov (2001): "Marginal Likelihood from the Metropolis-Hastings Output," *Journal of the American Statistical Association*, **96**(453), 270-281.

Clarida, Richard, Jordi Galí, Mark Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, **115**(1), 147-180.

Cooley, Thomas F., Stephen F. LeRoy, and Neil Raymon (1984): "Econometric Policy Evaluation: Note," *American Economic Review*, **74**(3), 467-470.

Del Negro, Marco and Frank Schorfheide (2002): "Priors from General Equilibrium Models for VARs," *Manuscript*, Department of Economics, University of Pennsylvania.

Erceg, Christopher J. and Andrew T. Levin (2001): "Imperfect Credibility and Inflation Persistence," *Finance and Economics Discussion Series*, **2001-45**, Federal Reserve Board.

Fernandez-Villaverde, Jesus and Juan Rubio-Ramirez (2002): "Estimating Nonlinear Dynamic Equilibrium Economies: A Likelihood Approach," *Manuscript*, Department of Economics, University of Pennsylvania.

Galí, Jordi and Mark Gertler (1999): "Inflation Dynamics: A Structural Econometric Analysis". *Journal of Monetary Economics*, **44**(2), pp. 195-222.

Geweke, John F. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, **18**, 1-126.

Hamilton, James D. (1994): *"Time Series Analysis,"* Princeton University Press.

Jeffreys, Harold (1961): *"Theory of Probability."* Oxford University Press, Oxford.

Kim, Chang-Jin (1994): "Dynamic Linear Models with Markov-Switching," *Journal of Econometrics*, **60**, 1-22.

Kim, Chang-Jin and Charles R. Nelson (1999): *"State-space Models with Regime Switching,"* MIT Press, Cambridge.

King, Robert G. (2000): "The New IS-LM Model: Language, Logic, and Limits". Federal Reserve Bank of Richmond *Economic Quarterly*, **86**, 45-103.

King, Robert G. and Alexander L. Wolman (1999): "What Should the Monetary Authority Do If Prices Are Sticky?" In: John B. Taylor (ed.): *Monetary Policy Rules*. University of Chicago Press.

Leeper, Eric M., and Tao Zha (2001): "Modest Policy Interventions," *Federal Reserve Bank of Atlanta Working Paper*, **99-22**.

Lubik, Thomas and Frank Schorfheide (2002): "Testing for Indeterminacy in Linear Rational Expectations Models," *Manuscript*, Department of Economics, University of Pennsylvania.

Lucas, Robert E. (1976): "Econometric Policy Evaluation: A Critique." In *Carnegie Rochester Conference Series on Public Policy*, 1, Karl Brunner and Allan H. Meltzer (eds.), North-Holland, Amsterdam, 104-130.

Sargent, Thomas J. (1999): *"The Conquest of American Inflation,"* Princeton University Press, Princeton.

Schorfheide, Frank (2000): "Loss Function-Based Evaluation of DSGE Models". *Journal of Applied Econometrics*, **15**, 645-670.

Schwarz, Gideon (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, **6**, 461-464.

Sims, Christopher A. (1980): "Macroeconomics and Reality," *Econometrica*, **48**, 1-48.

Sims, Christopher A. (1982): "Policy Analysis with Econometric Models," *Brookings Papers of Economic Activity*, **1**, 107-164.

Sims, Christopher A. (2000a): "Drift and Breaks in Monetary Policy," *Manuscript*, Department of Economics, Princeton University.

Sims, Christopher A. (2000b): "Solving Linear Rational Expectations Models". *Manuscript*, Department of Economics, Princeton University.

Sims, Christopher A. and Tao Zha (2002): "Macroeconomic Switching," *Manuscript*, Department of Economics, Princeton University.

Woodford, Michael (2000): "A Neo-Wicksellian Framework for the Analysis of Monetary Policy," *Manuscript*, Department of Economics Princeton University.

Table 1: PRIOR DISTRIBUTION

Name	Range	Density	Mean	90% Interval
γ	$I\!R$	Normal	0.50	[-0.32, 1.33]
$\ln \pi^*$	$I\!R^+$	Gamma	4.50	[2.43, 6.27]
$\ln \pi^*_L$	$I\!R^+$	Gamma	3.00	[1.35, 4.55]
$\ln \pi^*_H/\pi^*_L$	$I\!R^+$	Gamma	3.00	[1.35, 4.55]
$\ln r$	$I\!R^+$	Gamma	2.00	[0.47, 3.50]
τ	$[0,1]$	Beta	0.50	[0.41, 0.58]
κ	$I\!R^+$	Gamma	0.30	[0.13, 0.45]
ψ	$I\!R^+$	Gamma	1.62	[1.00, 2.22]
ρ_g	$[0,1)$	Beta	0.80	[0.65, 0.96]
ρ_z	$[0,1)$	Beta	0.30	[0.13, 0.46]
ρ_R	$[0,1)$	Beta	0.50	[0.18, 0.83]
σ_g	$I\!R^+$	InvGamma	1.25	[0.53, 1.99]
σ_z	$I\!R^+$	InvGamma	1.25	[0.53, 1.99]
$\sigma_{R,L}$	$I\!R^+$	InvGamma	0.63	[0.26, 1.00]
$\sigma_{R,H}$	$I\!R^+$	InvGamma	0.63	[0.26, 1.00]
ϕ_1	$[0,1)$	Beta	0.90	[0.83, 0.98]
ϕ_2	$[0,1)$	Beta	0.90	[0.83, 0.98]

Notes: The Inverse Gamma priors are of the form $p(\sigma|\nu,s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$, where $\nu = 4$ and s equals 1, 1, 0.5, and 0.5, respectively. The prior is truncated at the boundary of the determinacy region. Real interest rate r, target inflation rates π^*_L, π^*_H, π^* and the standard deviations of the policy shock $\sigma_{R,L}$ and $\sigma_{R,H}$ are annualized.

Table 2: PARAMETER ESTIMATION RESULTS

	No Switching		Full Information		Learning	
	Mean	Conf. Interval	Mean	Conf Interval	Mean	Conf Interval
γ	0.40	[0.26, 0.55]	0.39	[0.20, 0.59]	0.43	[0.23, 0.63]
$\ln \pi_L^*$	4.36	[3.62, 5.07]	2.83	[2.19, 3.49]	2.63	[1.87, 3.38]
$\ln \pi_H^*/\pi_L^*$			5.21	[3.94, 6.53]	4.78	[3.21, 6.28]
$\ln r$	2.10	[1.41, 2.82]	2.21	[1.68, 2.73]	2.20	[1.59, 2.82]
τ	0.01	[0.00, 0.01]	0.40	[0.31, 0.48]	0.39	[0.30, 0.47]
κ	0.37	[0.27, 0.47]	0.36	[0.26, 0.47]	0.37	[0.26, 0.47]
ψ	1.14	[1.00, 1.30]	1.68	[1.41, 1.96]	1.77	[1.32, 2.18]
ρ_g	0.98	[0.97, 1.00]	0.99	[0.98, 1.00]	0.98	[0.97, 1.00]
ρ_z	0.66	[0.62, 0.70]	0.78	[0.75, 0.82]	0.80	[0.76, 0.84]
ρ_R	0.82	[0.78, 0.87]	0.75	[0.70, 0.79]	0.76	[0.71, 0.81]
σ_g	1.20	[1.07, 1.32]	1.26	[1.13, 1.40]	1.25	[1.11, 1.39]
σ_z	0.28	[0.24, 0.32]	0.29	[0.25, 0.33]	0.29	[0.25, 0.33]
$\sigma_{R,L}$	0.99	[0.89, 1.08]	0.68	[0.59, 0.77]	0.62	[0.48, 0.74]
$\sigma_{R,H}$			1.71	[1.39, 2.02]	1.57	[1.31, 1.82]
ϕ_1			0.97	[0.96, 0.99]	0.96	[0.93, 0.99]
ϕ_2			0.95	[0.93, 0.98]	0.95	[0.92, 0.99]

Notes: The table reports posterior means and 90 percent confidence intervals (in brackets). The posterior summary statistics are calculated from the output of the posterior simulator. Real interest rate r, target inflation rates π_L^*, π_H^*, π^* and the standard deviations of the policy shock $\sigma_{R,L}$ and $\sigma_{R,H}$ are annualized.

Table 3: Data Densities and Bayes Factors

Specification	Log Data Density	Bayes Factor
Learning (Benchmark)	-806.80	1.00
Full Information	-801.05	314
No Regime Switching	-825.62	6E-9
Learning, $\sigma_{R,L} = \sigma_{R,H}$	- 833.21	3E-12
Full Information, $\sigma_{R,L} = \sigma_{R,H}$	-824.68	2E-08
Learning, hybrid Phillips Curve	- 807.91	0.33
Full Information, hybrid Phillips Curve	- 802.34	86.4
Learning, output growth and inflation rule	-807.05	0.78
Full Information, output growth and inflation rule	-782.06	5E10
VAR, Minnesota Prior	-717.10	9E38

Notes: The table reports log data densities $\ln p(Y^T|\mathcal{M})$ and Bayes factors relative to the benchmark Bayesian learning specification. The Bayes factor of $\mathcal{M}_i$ versus $\mathcal{M}_j$ can be obtained from $\exp[\ln p(Y^T|\mathcal{M}_i) - \ln p(Y^T|\mathcal{M}_j)]$.

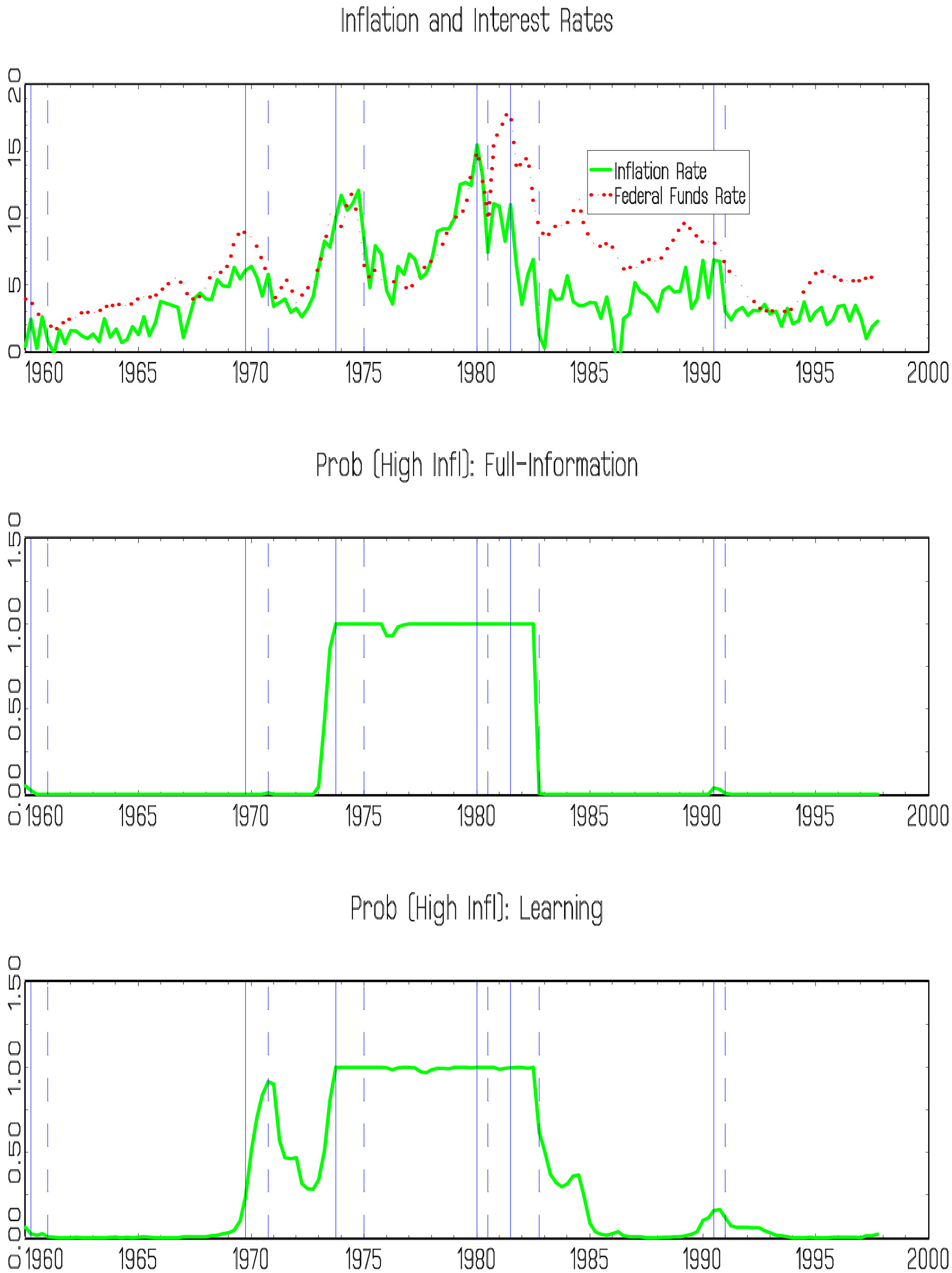

Figure 1: Regime Probabilities

Notes: Figure depicts posterior expected value of the monetary policy regimes for the 'full-information' and the 'learning' specification. As reference we also plot the inflation (solid line) and Federal Funds (doted line) rates.

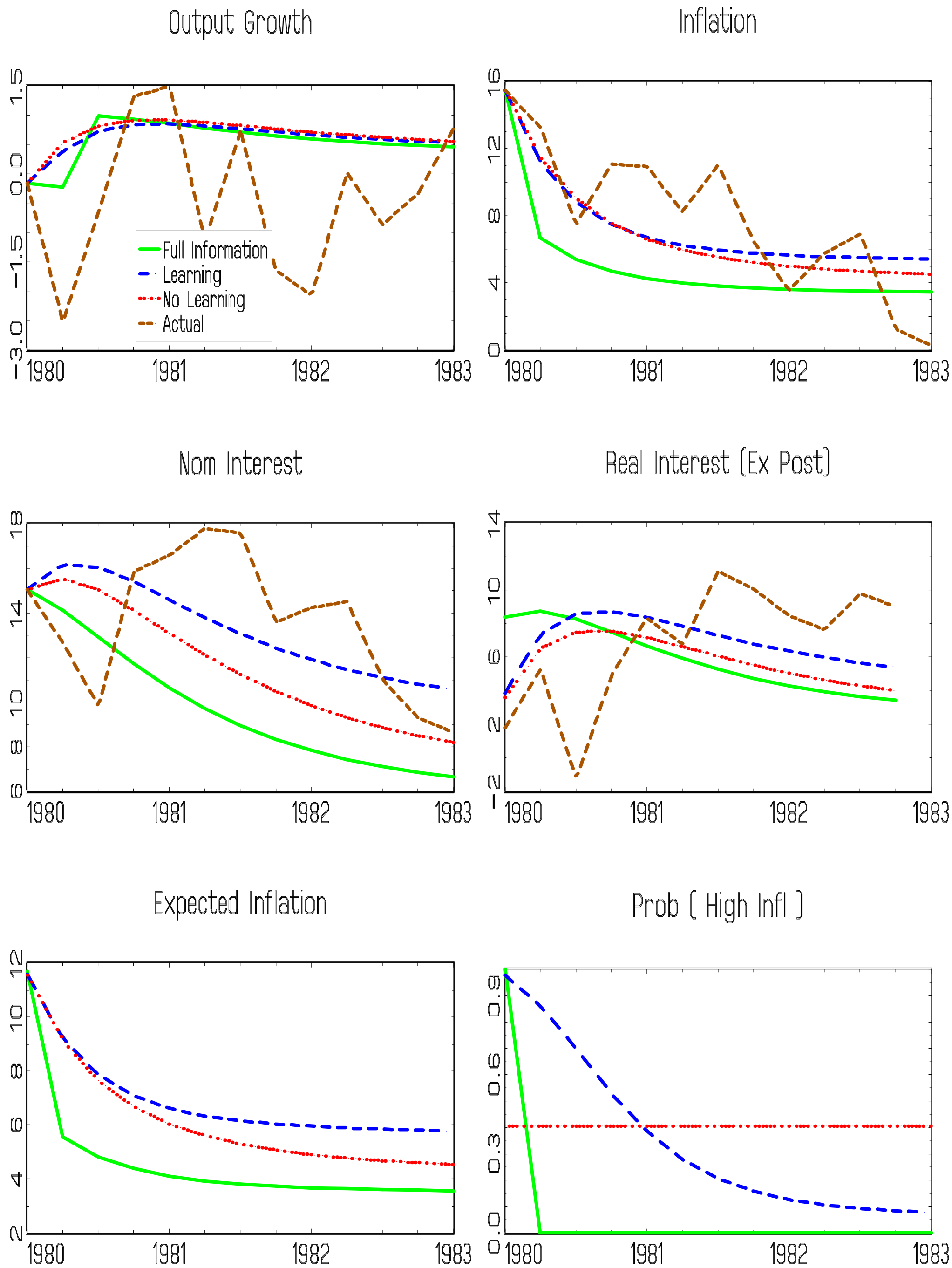

Figure 2: Disinflation Scenarios

Notes: Figure depicts posterior expected disinflation trajectories for the full information (solid line), learning (long dashes), no learning (dotted), specifications together with the actual values (short dashes) of output growth, inflation, and interest rates.

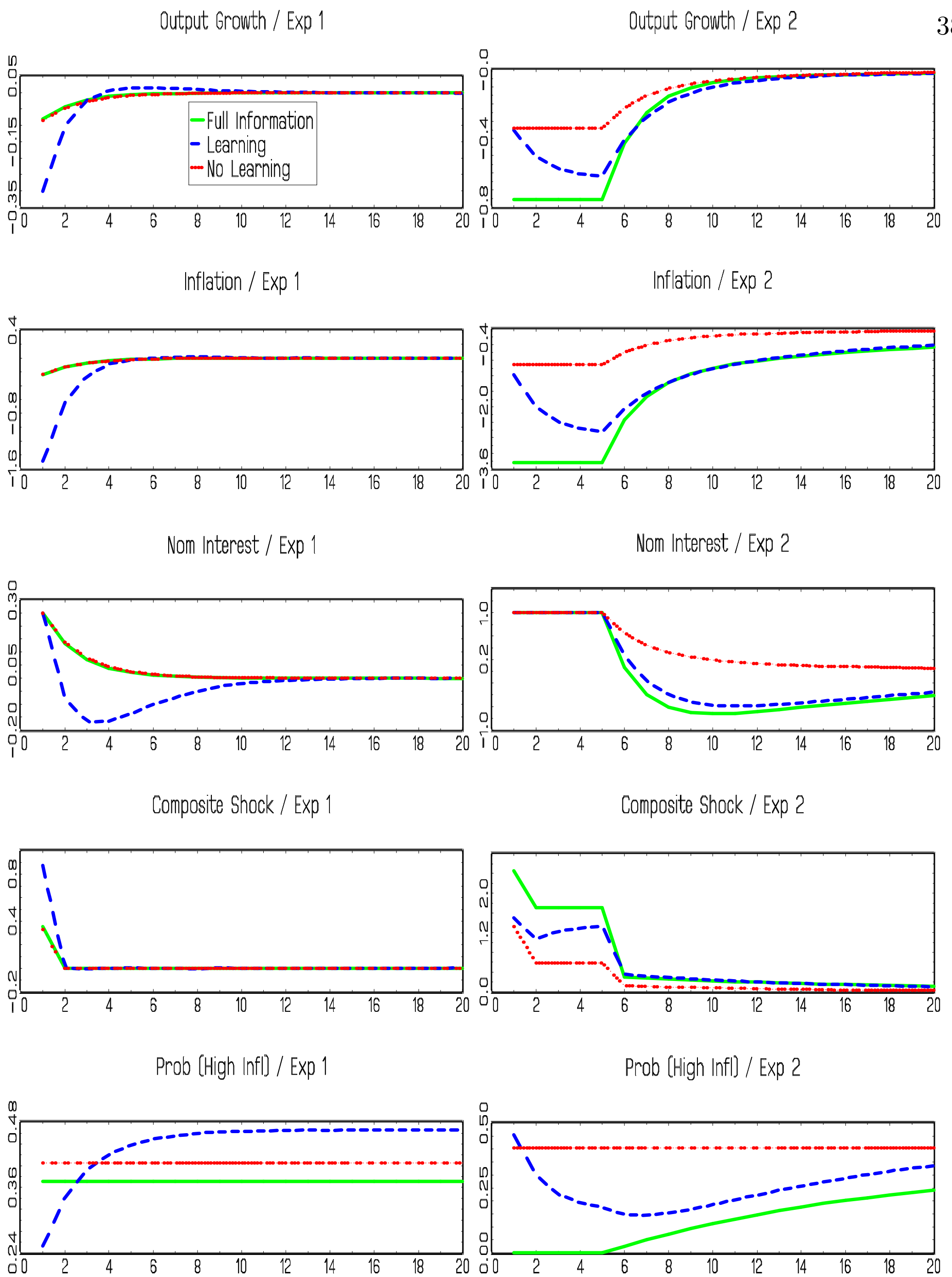

Figure 3: Effect of Policy Interventions

Notes: Figure depicts posterior mean responses: full information (solid line), learning (long dashes), no learning (dotted). Experiment 1: one-period intervention that raises the interest rate by 25 basis points. Experiment 2: five-period intervention that raises the interest rate by 100 basis points.

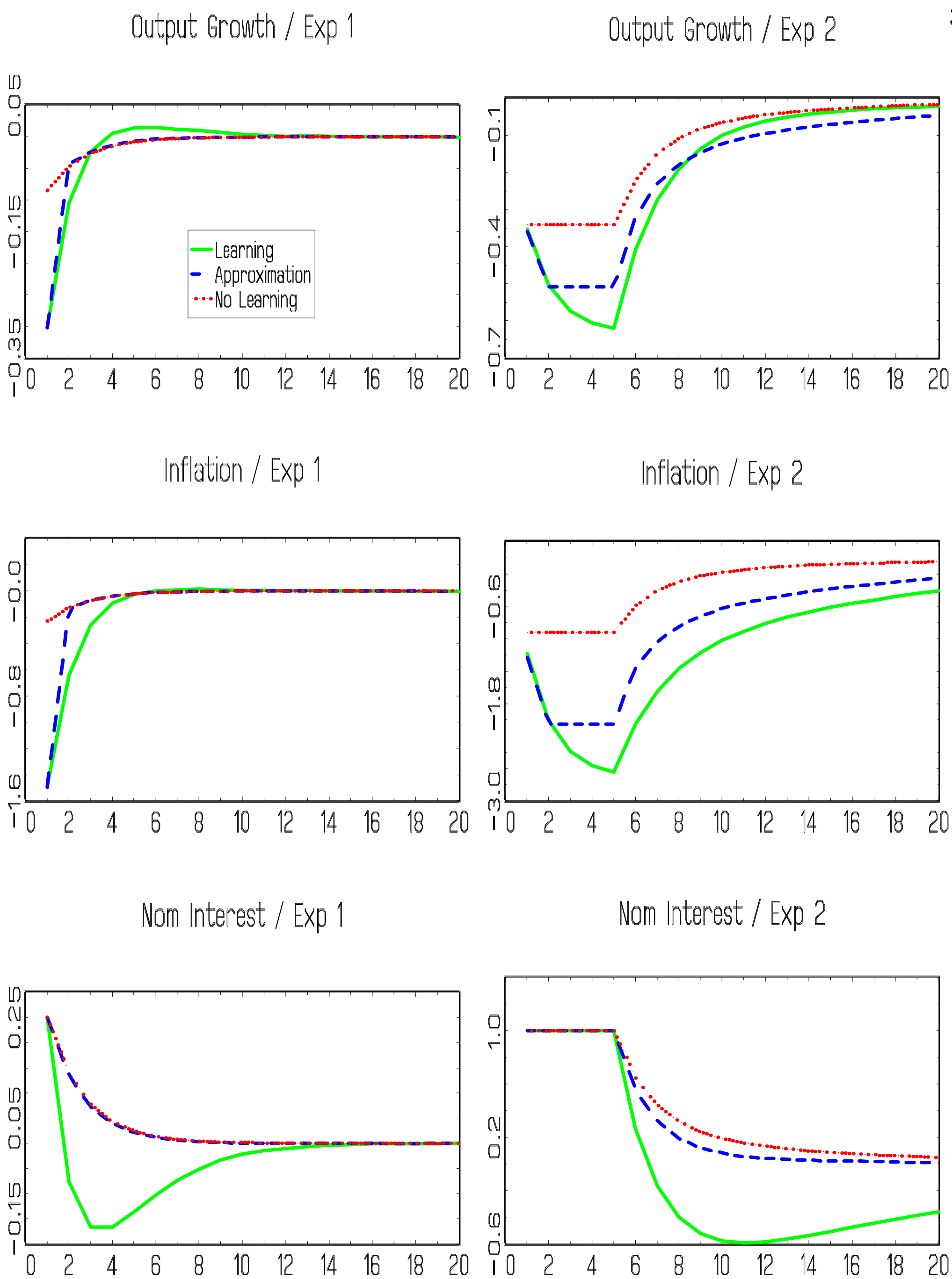

Figure 4: Approximating the 'Learning' Responses

Notes: Figure depicts posterior mean responses: learning (solid line), approximation (long dashes), no learning (dotted). Experiment 1: one-period intervention that raises the interest rate by 25 basis points. Experiment 2: five-period intervention that raises the interest rate by 100 basis points.

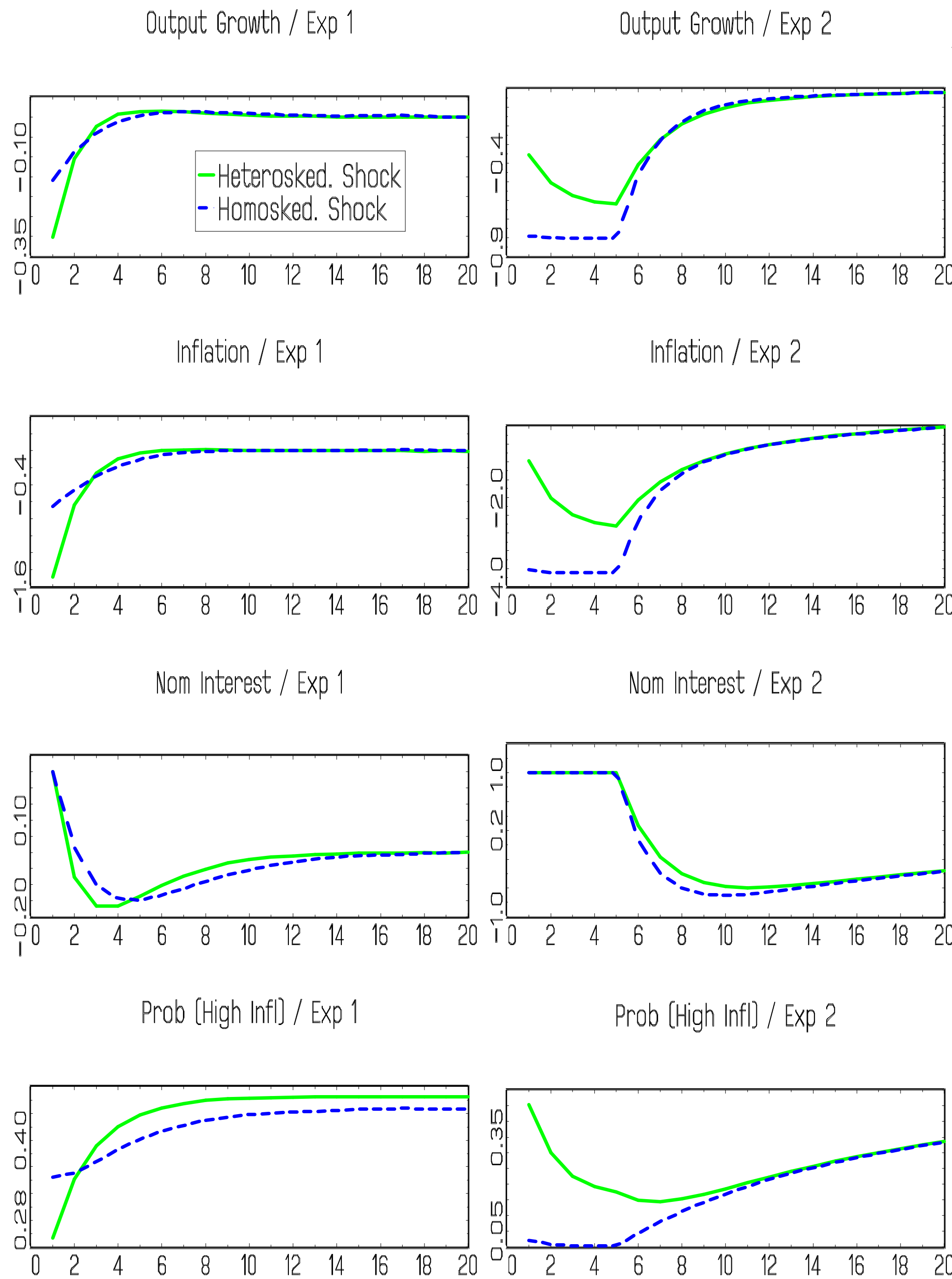

Figure 5: Effect of Regime-dependent Policy Shock Variances

Notes: Figure depicts posterior mean responses: learning under heteroskedastic policy shocks (solid line), learning under homoskedastic policy shocks approximation (long dashes). Experiment 1: one-period intervention that lowers the interest rate by 25 basis points. Experiment 2: five-period intervention that lowers the interest rate by 100 basis points.