MACROECONOMIC SWITCHING

CHRISTOPHER A. SIMS AND TAO ZHA

ABSTRACT. We discuss the results of fitting a 6-variable structural VAR in which we allow for certain types of parameter variation over time. Allowing structural equation variances to change over time is extremely important in improving fit. Allowing the coefficients that define the model's dynamics to change is less important to improving fit, though models with changing parameters are consistent with the data. We pay special attention to a version of the model that allows the monetary policy rule, but not other parts of the model, to show changing coefficients. Results from this model fit some aspects of conventional wisdom about changes in monetary policy over time, but imply that the changes in policy have been more subtle than dramatic. We construct counterfactual histories for the early 1980's, suppressing the "Volcker regime" in monetary policy. We find a steadier decline in inflation and a smaller recession earlier in the period, but slower growth later, than actually occurred.

1. INTRODUCTION

This paper aims to contribute to a recently active line of research that analyzes the evolution of monetary policy behavior and its potential effects on the economy. Some writers (Cogley and Sargent, 2001; Clarida, Galí, and Gertler, 2000, e.g.), and this may reflect a majority view, believe that the systematic component of US monetary policy has changed substantially since 1950, and that it is this change that explains the rise and subsequent decline in US inflation over the last 4 decades. However much of the most careful statistical analysis in this area (Hanson, 2001; Leeper and Zha, 2001a; Sims, 1999) has found little clear evidence of changed policy. In Hanson (2001) and Leeper and Zha (2001a) it has been found to be important to use a multiple-equation framework in studying these issues. There is some evidence of shift between interest-smoothing and short-run monetary aggregate targeting, at least during the early 1980's, and single-equation "reaction function" studies, even those that use instrumental variables, cannot allow for the simultaneity with money demand that arises in this case. On the other hand, Sims (1999) argues that time-varying residual variances are the most important instability in the US interest rate time series, to the extent that inference that ignores heteroskedasticity is likely to be quite misleading. Previous research with multiple equation methods has not allowed for time varying variances.

Date: February 2002.

Key words and phrases. Simultaneity, identification, dynamic multiple equations, time-varying parameters.

Email: `sims@princeton.edu` and `tzha@mindspring.com`. We thank Dan Waggoner for many helpful discussions. The views expressed herein do not necessarily reflect those of the Federal Reserve Bank of Atlanta or the Federal Reserve System.

There is of course a reason that previous research has not combined time varying variances with structural multiple-equation modeling. The resulting models are complex, and push the limits of what our computers and analytical capacity can handle. The results we report in this draft of the paper are still incomplete in some respects, because of the extended computing times required by some of our analysis. Additional work on the programs can probably speed them up substantially.

Our model allows the economy to switch among a finite number of states. In a version of the model that fits reasonably well, the variances of all equation disturbances change between states, but the coefficients change only in the monetary policy reaction function. We find that 3 states do a good job of picking up the time variation in policy. One of the states occurs mainly during 1980-82, a period when the stated policy of the Federal Reserve system, under Paul Volcker's chairmanship, was to focus on controlling reserves, not on smoothing interest rates. The estimates suggest that this was indeed a period in which policy targeted monetary aggregates, not interest rates. However this regime is estimated to be almost entirely confined to the 80-82 period. It does not represent a permanent shift to a new type of behavior. A second regime is estimated to prevail during most of the rest of the period we study, but is more completely dominant in the post-1982 sub-period. However, estimated differences among regimes are modest. None of the policy regimes are unstable, in the sense of moving interest rates less than 1-for-1 with inflation in the long run.

2. The Model

Our model is described by nonlinear stochastic dynamic simultaneous equations of the general form:

$$y_t' A_0(s_t) = x_t' A_+(s_t) + \varepsilon_t', t = 1, \ldots, T, \tag{1}$$

$$\Pr(s_t = i \mid s_{t-1} = k) = p_{ik}, i, k = 1, \ldots, h, \tag{2}$$

where s is an unobserved state, y is an $n \times 1$ vector of endogenous variables, x is an $m \times 1$ vector of exogenous and lagged endogenous variables, A_0 is an $n \times n$ matrix of parameters, A_+ is an $m \times n$ matrix of parameters, T is a sample size, and h is the total number of states. The transition matrix $P = [p_{ik}]$ is assumed to be irreducible and ergodic. This class of nonlinear models was introduced by Hamilton 1989; what is new here is our application to a simultaneous equation framework. We treat as given the initial lagged values of endogenous variables $Y_0 = \{y_{1-\ell}, \ldots, y_0\}$ where ℓ is the lag length in (1) and we do not introduce additional notation for these variables in the following analysis. For $t = 1, \ldots, T$, denote

$$Y_t = \{y_1, \ldots, y_t\}.$$

Structural disturbances are assumed to have the distribution:

$$\varepsilon_t \mid Y_{t-1} \sim N\left(0, \underset{n \times n}{I}\right),$$

where $N(a, b)$ refers to the normal distribution with mean a and covariance matrix b.

The reduced-form parameter matrix, denoted by $B(s_t)$, is:

$$B(s_t) = A_+(s_t) A_0^{-1}(s_t). \tag{3}$$

Because both $A_0(s_t)$ and $A_+(s_t)$ vary with states, the reduced form of (1) has both time-varying parameters and heteroscedastic disturbances.

If we did not allow simultaneity (i.e., $A_0(s_t)$ is assumed to be recursive) and let all parameters vary across states, estimation of the model would be relatively straightforward, because A_0 and A_+ in each given state can be estimated independent of the parameters in other states. But with such an unrestricted form for the time variation, if the system of equations is large or the lag length is long, the number of free parameters in the model becomes impractically large. For a typical monthly model with 13 lags on 6 endogenous variables, for example, the number of parameters in $A_+(s_t)$ is of order 468. But given the post-war macroeconomic data, it is not uncommon to have some states lasting for only a few years and thus the number of associated observations is far less than 468. It is therefore essential to simplify the model by postulating restricted forms for the time variation in its parameters.

Our approach is to begin by rewriting A_+ as

$$\underset{m\times n}{A_+(s_t)} = \underset{m\times n}{D(s_t)} + \underset{m\times n}{\overline{S}}\,\underset{n\times n}{A_0(s_t)}. \tag{4}$$

where

$$\overline{S} = \begin{bmatrix} \underset{n\times n}{I} \\ \underset{(m-n)\times n}{\mathbf{0}} \end{bmatrix}.$$

Then by specifying a prior and model for D_t that makes its mean zero, we keep this prior centered on the same random walk model for the reduced form that has worked well as a prior mean in previous Bayesian VAR models. The approach of specifying a prior in terms of this A_0 and D matches the approach of Sims and Zha 1998a. Note also that, as can be seen from (3) and (4), this form of prior tends to imply that greater persistence (in the sense of tighter concentration of the prior on the random walk) is associated with smaller disturbance variances. This is reasonable, as it is consistent with the idea that beliefs about the unconditional variance of the data are *not* highly correlated with beliefs about the degree of persistence in the data.

The class of models we consider includes three types of *restricted* time variations for both $A_0(s_t)$ and $D(s_t)$. Each column of these matrices (i.e., the coefficients of each equation in the system) can be either constant, changing by a scale factor over time, or changing freely over time. When the change over time is "free", in fact some individual coefficients may be constrained to be zero. More specifically, in this paper we consider these three cases:

$$A_{0,j}(s_t), D_t = \begin{cases} \overline{A}_{0,j}, \overline{D}_j & \text{Case I} \\ \overline{A}_{0,j}\lambda_j(s_t), \overline{D}_j\lambda_j(s_t) & \text{Case II} \\ A_{0,j}(s_t), \overline{D}_j\lambda_j(s_t) & \text{Case III} \end{cases}, \tag{5}$$

where the subscript j denotes the j^{th} column (equation) of the corresponding matrix.

The specific models we discuss are[1]

[1] Our software can handle other combinations of the cases as well.

TABLE 1. Identifying restrictions on $A_0(s_t)$

Variable (below)	**Sector (right)**	Inf	Fed	MD	Prod	Prod	Prod
Pcom		X					
M		X	X	X			
R		X	X	X			
y		X		X	X	X	X
P		X		X		X	X
U		X					X

Nothing: Case I for all equations;
Variances only: Case II for all equations;
Monetary policy: Case II for all equations, except Case III for the monetary policy reaction function; and
Everything: Case III for all equations.

We use monthly US data from 1959:1–2001:6. The model has 13 lags and includes constant terms and 6 commonly-used endogenous variables: a commodity price index (Pcom), M2 (M), the federal funds rate (R), interpolated monthly real GDP (y), the consumer price index (P), and the unemployment rate (U). All variables are expressed in natural logs except for the federal funds rate and the unemployment rate which are expressed in percent. The identification of monetary policy, following Leeper and Zha (2001b), is described in Table 1. The X's in Table 1 indicate the unrestricted parameters in $A_0(s_t)$ and the blank spaces indicate the parameters that are restricted to be zero. The "Fed" column in Table 1 represents the Federal Reserve contemporaneous behavior; the "Inf" column describes the information sector (the commodity market); the "MD" represents the money demand equation; and the block consisting of the last three columns represents the production sector, whose variables are arbitrarily ordered in an upper triangular form.[2]

There are two major factors that make the estimation and inference of our model a difficult task. One factor is possible simultaneous relationships in the structural matrix $A_0(s_t)$. The other factor is the types of restricted time variations specified in (5). Appendix sections detail a Bayesian estimation method used for our structural model.

3. RESULTS

Our most detailed discussion concerns estimates of the "Monetary policy" model, which is the one in which all equations other than the monetary policy reaction function have constant coefficients and time-varying disturbance variances, while for the monetary policy equation both a scale factor λ_t on D_j and all the coefficients in A_{0j} are allowed to vary over time.

[2]While we provide no discussion here of why delays in reaction of the private sector to financial variables might be plausible, explanations of inertia, and examination of its effects, are common in the recent literature (Christiano, Eichenbaum, and Evans, 2001; Edge, 2000; Sims, 2001a, 1998, e.g.).

TABLE 2. Contemporaneous coefficient matrices, $A_0(1)$

	Financial	M Policy	M demand	Private y	Private P	Private U
Pcom	70.72	0.00	0.00	0.00	0.00	0.00
M2	27.97	−545.82	−359.16	0.00	0.00	0.00
R	−36.05	238.83	−643.94	0.00	0.00	0.00
y	−6.45	0.00	35.07	303.68	−23.06	60.05
P	−41.52	0.00	50.36	0.00	−623.98	10.96
U	60.61	0.00	0.00	0.00	0.00	721.96

TABLE 3. Contemporaneous coefficient matrices, $A_0(2)$

	Financial	M Policy	M demand	Private y	Private P	Private U
Pcom	44.53	0.00	0.00	0.00	0.00	0.00
M2	17.61	−314.34	−141.47	0.00	0.00	0.00
R	−22.70	74.63	−253.64	0.00	0.00	0.00
y	−4.06	0.00	13.81	217.70	−20.29	51.31
P	−26.14	0.00	19.84	0.00	−549.10	9.37
U	38.16	0.00	0.00	0.00	0.00	616.86

TABLE 4. Contemporaneous coefficient matrices, $A_0(3)$

	Financial	M Policy	M demand	Private y	Private P	Private U
Pcom	39.36	0.00	0.00	0.00	0.00	0.00
M2	15.57	−509.75	−38.60	0.00	0.00	0.00
R	−20.06	−8.36	−69.21	0.00	0.00	0.00
y	−3.59	0.00	3.77	199.20	−11.20	37.77
P	−23.11	0.00	5.41	0.00	−303.05	6.90
U	33.73	0.00	0.00	0.00	0.00	454.15

Tables 2-4 show the variation in A_0 across states in this model. The changes between states for the monetary policy equation are nontrivial. If the policy equation were normalized on the interest rate R as is conventionally done, the standard deviation of the residual would be 28 times bigger in state 3 than in state 1 and 9 times bigger in state 3 than state 1. But this is not a simple rescaling of the equation; as the R coefficient goes down, the M coefficient goes up, consistent with short-run reserves targeting, which is what the Fed said it was doing during 1979 – 1982.

Figure 1 displays the probabilities of the states for our model, plotted against a background showing the time path of the Federal Funds rate. Clearly state 3 shows up in rare circumstances, mainly in the 1979-82 reserve-targeting period, but also during a high-interest-rate period in the early 70's. State 2 differs from state 1 in having lower precision

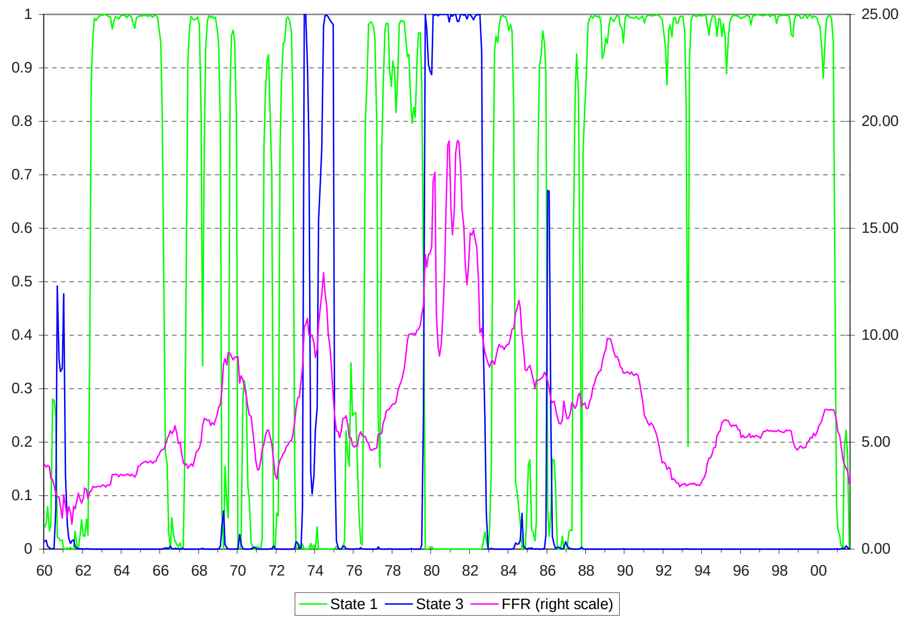

FIGURE 1. Probabilities of states

TABLE 5. .90 probability bands of the short run policy coefficients

	state 1	state 2	state 3
M2	$(-416.4, -51.7)$	$(-317.4, -83.74)$	$(-510.1, -356.3)$
R	$(214.4, 548.5)$	$(102.5, 307.3)$	$(-18.6, 46.8)$

(higher disturbance variance) in the interest rate. It was more common before 1987 than it has been since, though it seems to be coming back in the last few months of the sample (early 2001). One might argue that this is a "regime shift", but the changes between these two states seem modest, and the shift does not seem to have been permanent.

The .90 probability intervals for short run (contemporaneous) coefficients of the monetary policy equation are reported in Table 5. The posterior estimates of the M coefficient for state 1 and R coefficient for state 2 are outside the .90 probability intervals, reflecting thin, curved ridges of the joint posterior distribution and possibly our normalization rule for labeling the states (Waggoner and Zha 2000 and 2001). Table 5 shows that the signs of all coefficients except that of R in state 3 are estimated as "significant" — that is, as having determinate sign with high probability.

The long run policy responses of R to all other five variables, along with .90 probability intervals, are reported in Table 6. These long run responses are calculated simply as $\sum \alpha_k / \sum \gamma_k$, where α_k is the coefficient on the k'th lag of the "right-hand-side" variable in

TABLE 6. Long run policy responses of R to

	state		
	1	2	3
Δ Pcom	0.09		
.90 prob interval	$(0.06, 0.69)$		
Δ M2	1.50	0.82	4.60
.90 prob interval	$(0.50, 115.49)$	$(-0.53, 1.90)$	$(1.43, 8.37)$
Δ y	-1.07		
.90 prob interval	$(-1.41, 1.26)$		
Δ P	0.39		
.90 prob interval	$(-1.50, 0.97)$		
U	-0.55		
.90 prob interval	$(-2.06, 0.94)$		

the policy reaction function and γ_k is the coefficient on the k'th lag of R in the policy equation. Note that because of (4), the sum of current and lagged coefficients on R moves in proportion to $\lambda_j(s_t)$, as does the sum of coefficients on lagged U, for example. Because the sums of coefficients on lagged values of Pcom, P, M2 and y are nearly zero, it makes more sense with these variables to give the responses to the differenced variables, calculated as

$$\frac{\sum_k = 0^\ell \sum_{i=0}^k \alpha_k}{\sum_{k=0}^\ell \gamma_j} .$$

The differences in all other log variables, such as Pcom and M, are annualized to match the annual rate of R. Because of the relation and the identifying restrictions imposed on the policy equation of $A_0(s_t)$, it can be easily seen that long run responses to ΔPcom, Δy, ΔP, and U are the same across all states, represented by blank spaces in Table 6.[3] Differences across states show up in the policy response to ΔM2. The posterior estimate of the policy response to output growth (Δy) has a "wrong" sign but is very imprecise according to the .90 probability interval. The policy responses to inflation (ΔP) and the unemployment rate (U) have correct signs but they are imprecisely estimated as well by the .90 intervals. The interest rate response to changes in commodity prices (ΔPcom) is positive and sharply estimated.

The interest rate response to money growth (ΔM) in state 1 is greater than 1, something like the Taylor rule. But the distribution of this response has a long fat tail towards infinity and the probability band is very large. In contrast, the same response in state 3 is more sharply estimated and the response in state 3 at the peak of the distribution is at least four times larger than in state 1. In state 2, the Federal Reserve responds more weakly to money growth. The response is insignificant by the .90 probability interval. By the .68 probability

[3]Our model is composed of multiple equations. Thus, long run *reduced-form* responses of the interest rate to other variables will be different across states.

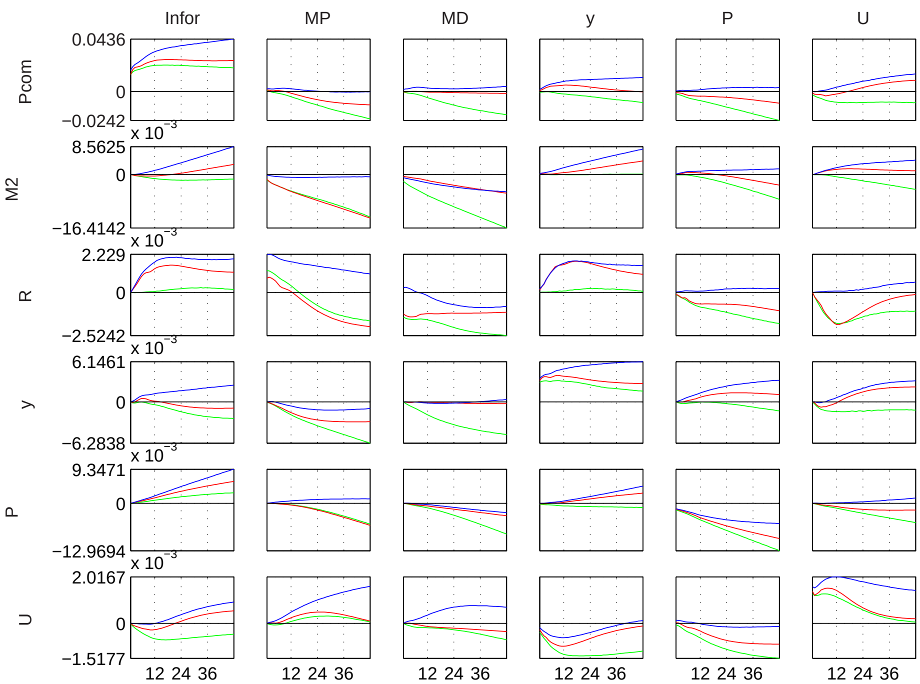

FIGURE 2. Impulse responses, state 1

interval (not reported in Table 6, the response is positive but less than 1. This does not mean that state 2 is likely to represent an unstable regime, however, because stability is determined by the sum of long run responses on all nominal variables in the system, and this is above one as a point estimate, albeit apparently with uncertainty extending from negative values to far above one.[4]

We report impulse responses in different states, along with .68 probability bands, in Figure 2-4. The responses to monetary policy shocks accord well with the findings in the recent literature, confirming the validity of our identification.

4. COUNTERFACTUAL HISTORY

It is difficult from looking at coefficients, impulse responses, and long run responses to form a clear idea of the quantitative significance of the differences between estimated regimes. One way to make this clearer is to rerun economic history, replacing the 1979-82 Volcker reserve-targeting regime (state 3) with the state 1 regime, the one that immediately preceded the reserve-targeting regime and that has prevailed through most of the 90's. Would this change in policy have greatly affected outcomes? Would inflation have taken much longer to end?

We examine the historical period 1979:10-1987:7. The counterfactual exercise can be done in our model as follows. Given the data Y_T, we draw (θ, S_T) jointly. For each draw,

[4]The fact that our apparently flexible specification actually constrains the policy rule long run responses to change only for ΔM2 is restrictive, and may limit the persuasiveness of our claim that policy changes have not been large. We intend to look at other specifications that relax this constraint.

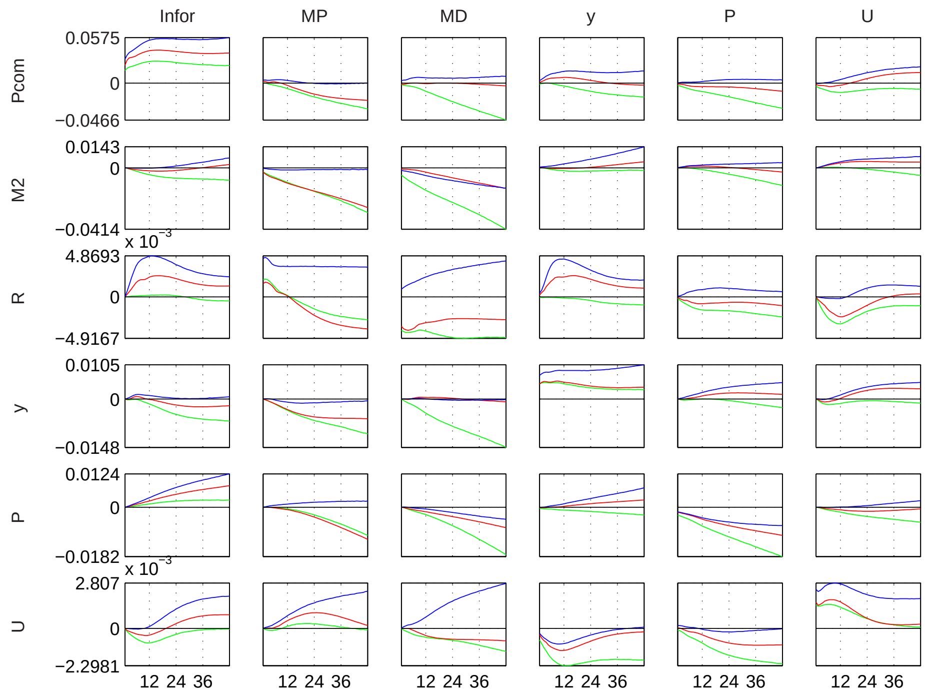

FIGURE 3. Impulse responses, state 2

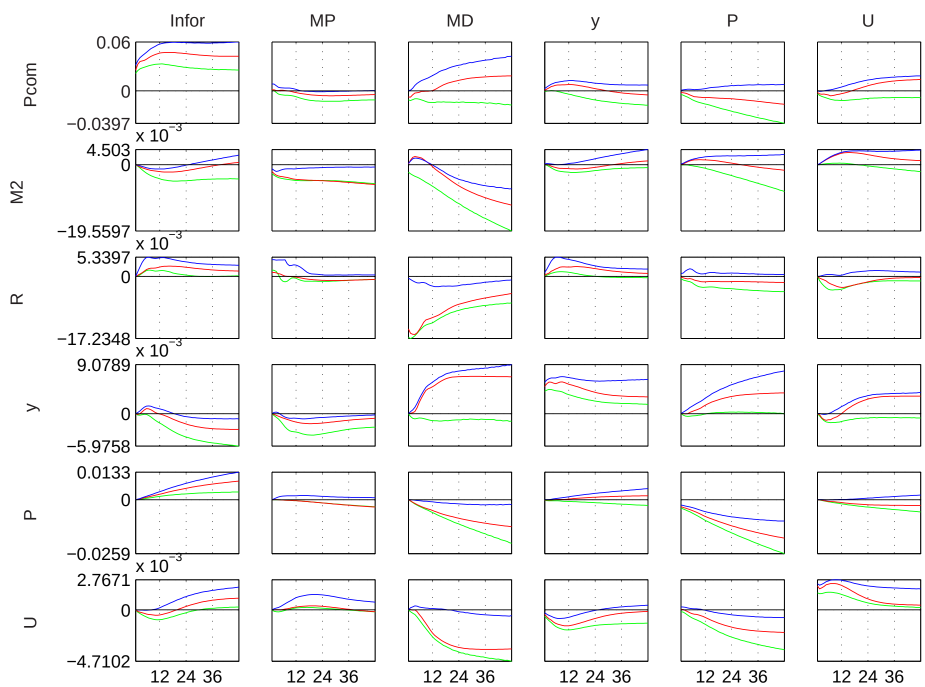

FIGURE 4. Impulse responses, state 3

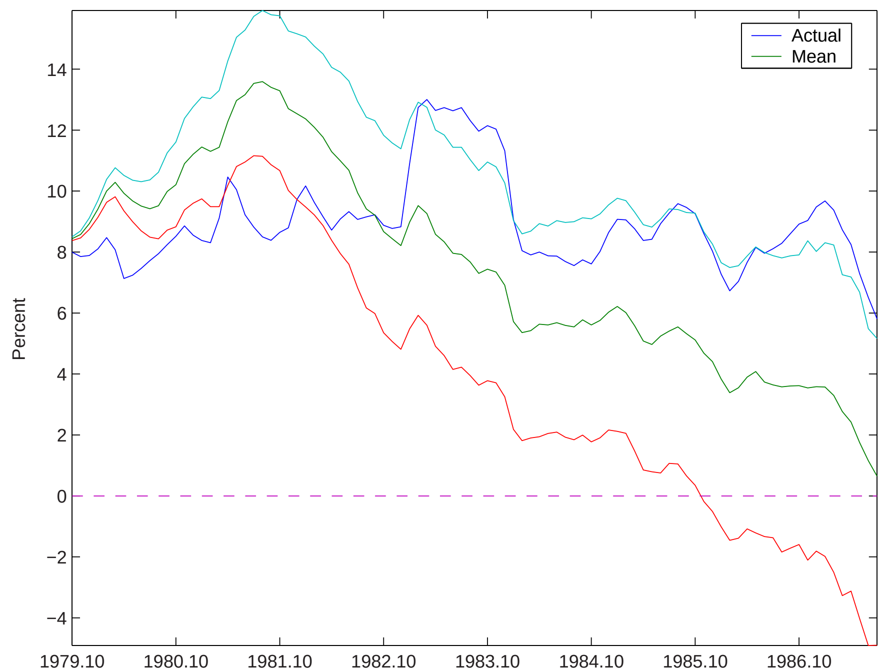

FIGURE 5. Counterfactual paths conditional on state 1: M2

we back out a sequence of unit-variance structural shocks $(\tilde{\varepsilon}_1, \ldots, \tilde{\varepsilon}_T)$. Conditional on the data up to September 1979 and the model parameters in state 1, we generate a one-step forecast with the structural shock set to $\tilde{\varepsilon}_{1979:10}$ and then a two-step forecast with the structural shock set to $\tilde{\varepsilon}_{1979:11}$. We continue this iteration until 1987:7 is reached. This counterfactual path represents one that would have taken place when the regime had been in state 1. For thousands of draws, we compute the mean path as well as .68 probability bands. The results are reported in Figures 5-9 where the time label 1979.1 means 1979:10 and so on.

Clearly, the counterfactual funds rate path is much smoother than actual data: the funds rate would have been lower from late 1979 to the end of 1981 but higher after the mid 1982, coupled with a higher growth rate of money initially but lower growth later (Figures 5 and 6). Consequently, the counterfactual inflation path would have come down as steadily as actual data (Figure 9). But the tradeoff would have been equally severe except the timing would have been postponed: from the end of 1984 on, the error bands imply that we would have had a much lower growth rate of GDP or even recessions (Figure 7) and a much higher level of the unemployment rate (Figure 9).

5. MODEL FIT

Table 7 provides measures of fit for the model used so far by comparing it to other models. The "4 states" line refers to the extension of our Monetary-policy-only model to four states. The upper panel of the table shows nearly-conventional fit criteria — likelihood,

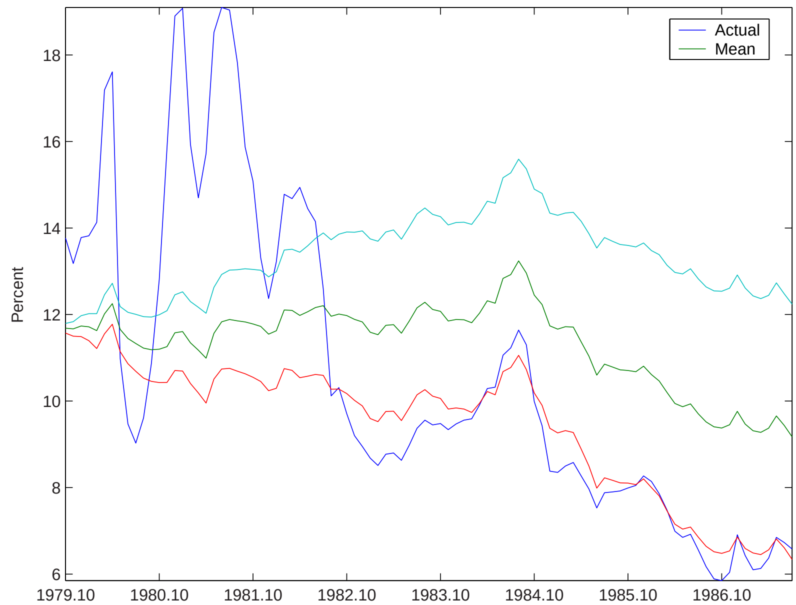

FIGURE 6. Counterfactual paths conditional on state 1: FFR

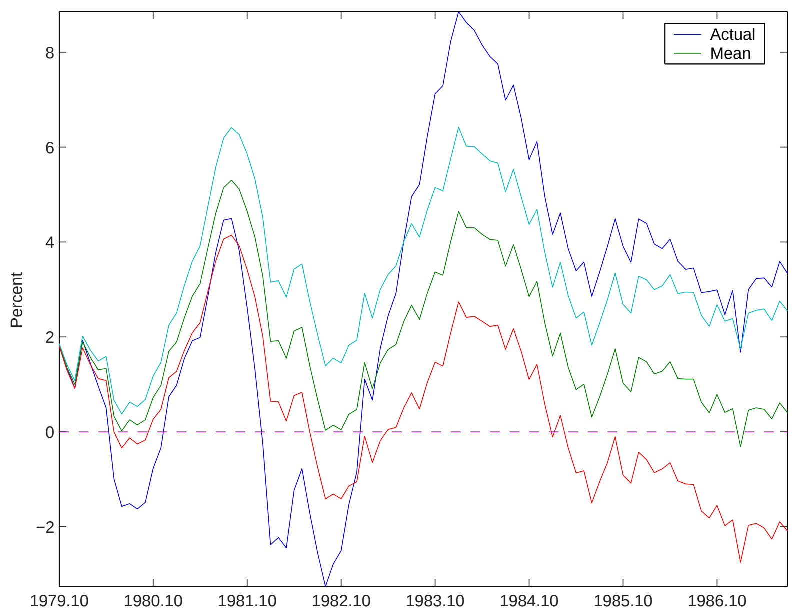

FIGURE 7. Counterfactual paths conditional on state 1: y

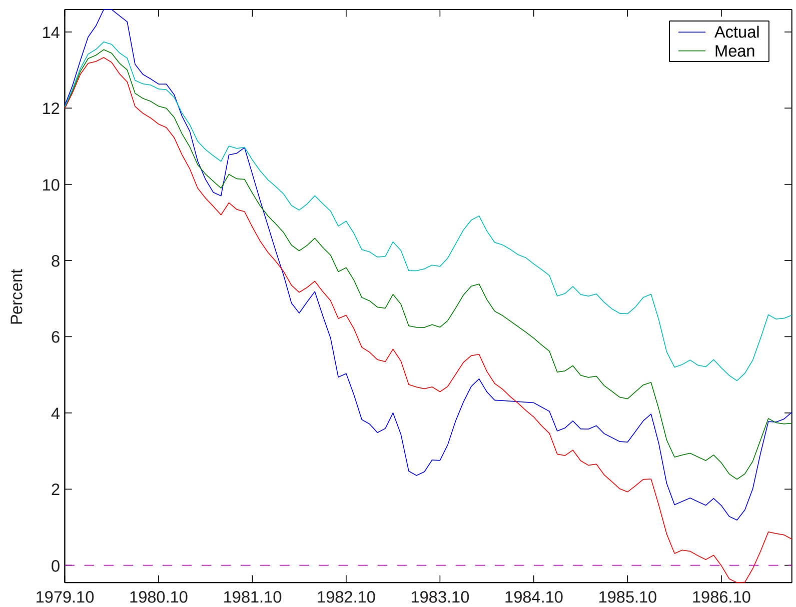

FIGURE 8. Counterfactual paths conditional on state 1: P

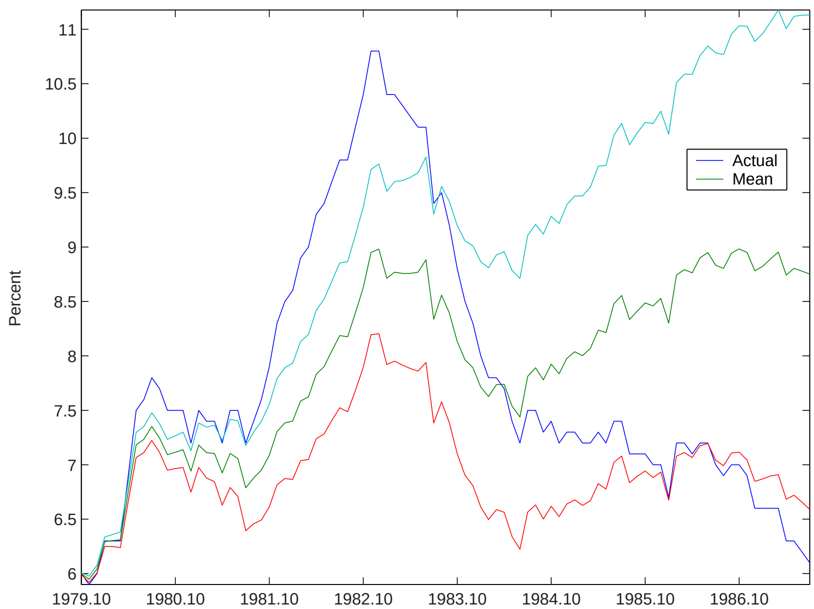

FIGURE 9. Counterfactual paths conditional on state 1: U

TABLE 7. Measures of Fit

Based on likelihood alone

	$\log f(Y_T\|\hat{\theta})$	DF	Akaike	Schwarz
Nothing	12852		12852	12852
Variances only	13244	12	13232	13207
M policy	13248	16	13232	13198
Everything	13250	48	13202	13100
4 states M policy	13279	24	13255	13204

Based on posterior density

	$\log(f(Y_T\|\hat{\theta})\cdot\pi(\hat{\theta}))$	$\pi(\hat{\theta})$	DF	Akaike	Schwarz	$\pi(Y_T)$
Nothing	12852	-1915.5		10937	10937	
Variances only	13244	-2095.5	12	11137	11111	12432
M policy	13248	-2109.3	16	11123	11089	12433
Everything	13250	-2264.7	48	10937	10836	12476
4 states M policy	13279	-2254.2	24	11001	10950	12464

the Schwarz criterion, and the Akaike criterion. The only unconventional aspect here is that they are evaluated at the peak of the posterior pdf, not at the likelihood peak. To the extent that the likelihood dominates the prior, they can be interpreted like usual SC and AC statistics. With as many free parameters as are present here, the Akaike criterion is probably unreliable. The Schwarz criterion favors the Variances-only model, but not by a great margin over the 4-states model. The Monetary-policy model to which we have paid the most attention comes in third.

In the second panel, we add to the likelihood values the log prior pdf values, to arrive at posterior pdf values. Here the usual argument for the Schwarz criterion's asymptotic validity still applies. The only change in the ranking of the models by the SC is that here "Everything" comes out even worse than "Nothing". The last column of this table is in principle the most reliable measure of fit. It measures the integrated posterior pdf over each model's parameter space, from which Bayes factors can be constructed. The results by this criterion are drastically different, with the Everything model coming out on top. However, we are still uncertain that the calculations underlying this column have converged, so its implications should not at this point be taken too seriously.

6. CONCLUSION

We have found more evidence in favor of stable dynamics with unstable disturbance variances than of clear changes in model dynamics. Nonetheless, our methods have found weak evidence of policy shifts that match intuition — targeting of monetary aggregates and abandonment of interest rate smoothing during 1979-82. The story that policy has changed drastically between the 60-78 period and the 83-2000 period does not seem to be borne out.

There are obvious caveats to these results. We need to finish the calculations of the marginal posterior model probabilities, and examine their sensitivity to choices of priors.

We need to examine a specification that would allow long run responses of policy to vary over time.

There are also clear opportunities to extend the results. For example, if we used a discretized version of an AR specification to parameterize P, the state transition matrix, we might be able to handle many more states and perhaps get a better fit. This approach has advantages over that of Cogley and Sargent (2001), who assume a continuously distributed state, because it could allow for occasional discontinuous shifts in regime as well as more frequent incremental changes.

APPENDIX A. PRIOR RESTRICTIONS

We consider two classes of additional prior restrictions commonly used in the literature. One class pertains to reference prior distributions for eliminating over-fitting problems; the other class pertains to non-stochastic linear restrictions on $A_0(s_t)$ and $D(s_t)$, such as commonly used exclusion restrictions, for achieving identification.

For $j = 1, \ldots, n$ and $k = 1, \ldots, h$, let $a_{0,j}(k)$ be the j^{th} column of $A_0(s_t)$ for $s_t = k$ and $d_j(k)$ be $D_j(s_t)$ for $s_t = k$ or $\overline{D}_j$ for all k. Prior distributions take the Gaussian form:

$$a_{0,j}(k) \sim \mathcal{N}(\mathbf{0}, H_{0j}),\ k = 1, \ldots, h,\ j = 1, \ldots, n; \tag{6}$$

$$d_j(k) \sim \mathcal{N}(\mathbf{0}, H_{+j}),\ k = 1, \ldots, h,\ j = 1, \ldots, n. \tag{7}$$

In addition, we follow Sims and Zha 1998a to incorporate into the model the two components of "dummy observations" prior components that express beliefs in unit roots and cointegration in macroeconomic series.[5]. In our application, we set H_{0j} and H_{+j} the same way as in Sims and Zha 1998a but scale them by the number of states (h) so that the Case I model in (5) coincides with the Bayesian VAR with constant parameters.

As shown in Waggoner and Zha 2000, linear restrictions imposed on $A_0(s_t)$ and $D(s_t)$ imply the following relationships:

$$\underset{nh\times 1}{a_j} = \underset{nh\times o_j}{U_j}\ \underset{o_j\times 1}{b_j},\ j = 1, \ldots, n; \tag{8}$$

$$\underset{mh\times 1}{d_j} = \underset{mh\times r_j}{V_j}\ \underset{r_j\times 1}{g_j},\ j = 1, \ldots, n; \tag{9}$$

where b_j and g_j are the free parameters "squeezed" out of a_j and d_j by the linear restrictions, o_j and r_j are the numbers of corresponding free parameters, columns of U_j are orthonormal vectors in the Euclidean space $\mathbb{R}^n$, columns of V_j are orthonormal vectors in $\mathbb{R}^m$, and

$$a_j = \begin{bmatrix} a_{0,j}(1) \\ \vdots \\ a_{0,j}(h) \end{bmatrix}, \quad d_j = \begin{bmatrix} d_j(1) \\ \vdots \\ d_j(h) \end{bmatrix}.$$

Combining (6) with (8) and (7) with (9) leads to the prior distributions for the free parameters b_j and g_j:

$$b_j \sim \mathcal{N}(\mathbf{0}, \overline{H}_{0j}), \tag{10}$$

$$g_j \sim \mathcal{N}(\mathbf{0}, \overline{H}_{+j}), \tag{11}$$

where

$$\overline{H}_{0j} = \left(U_j'(I \otimes H_{0j}^{-1})U_j\right)^{-1},$$

$$\overline{H}_{+j} = \left(V_j'(I \otimes H_{+j}^{-1})V_j\right)^{-1}.$$

[5] For detailed description of these dummy observations, see Sims and Zha 1998a

The prior distribution for $\lambda_j(k)$ is taken as $\lambda_j^2(k) \sim \Gamma(\alpha_\lambda, \beta_\lambda)$ in the Variances Only model, and $\lambda_j(k) \sim N(0, \sigma_\lambda^2)$ in the other models where λ appears. $\Gamma(\cdot)$ denotes the standard gamma distribution, with β_λ a scale factor (not an inverse scale factor as in the notation of many textbooks). The parameters β_λ and σ_λ are set at values so large that the prior distribution is essentially diffuse and has little influence on the posterior outcome.[6]

The prior of P takes a Dirichlet form as suggested by Chib 1996. For the k^{th} column of P, p_k, the prior density is

$$\pi(p_{1k}, \ldots, p_{hk}) = \mathcal{D}(\alpha_{1k}, \ldots, \alpha_{hk}) \propto p_{1k}^{\alpha_{1k}-1} \cdots p_{hk}^{\alpha_{hk}-1}, \; \alpha_{ik} > 1 \text{ for } i = 1, \ldots, h. \tag{12}$$

Although $\alpha_{ik} > 0$ is all one needs to have a proper Dirichlet, the condition $\alpha_{ik} > 1$ in (12) guarantees that the distribution of each p_{ik} has a well-defined unique mode. There are two steps in setting up a prior for p_k. First, the prior mode of p_{ik} is chosen to be υ_{ik} such that $\sum_{i=1}^h \upsilon_{ik} = 1$. Let the maximum of the prior mode vector υ_k correspond to the τ^{th} element. In the second step, set a prior variance on $\upsilon_{\tau k}$. Once υ_k and $\text{Var}(\upsilon_{\tau k})$ are given, we can solve for $\alpha_{\tau k}$ through a third order polynomial. Given $\alpha_{\tau k}$ and υ_k, we can solve for all other elements of the vector α_k through a system of $h-1$ linear equations. In our analysis, υ_{ik} is chosen to be evenly distributed across i for given k. The prior standard deviation of p_{ik} is chosen to be 0.235 for all i and k. The resulting prior distribution of P is very flat and has no influence on the posterior outcome.

Appendix B. Likelihood Function

Denote

$$\begin{aligned} b &= \{b_j,\ j = 1, \ldots, n\}; \\ \lambda &= \{\lambda_j(k),\ k = 1, \ldots, h,\ j = 1, \ldots, n\}; \\ g &= \{g_j,\ j = 1, \ldots, n\}; \\ p &= \{p_k,\ k = 1, \ldots, h\}; \\ \delta &= \{\lambda, g, p\}; \\ \theta &= \{b, \delta\}. \end{aligned}$$

The overall likelihood function $f(Y_T \mid \theta)$ can be obtained by integrating the conditional likelihood at each time t over s_t and recursively multiplying these conditional likelihood functions forward (Kim and Nelson 1999). Specifically,

$$f(Y_T \mid \theta) = \prod_{t=1}^{T} \left\{ \sum_{s_t=1}^{h} \left[f(Y_t \mid Y_{t-1}, s_t, \theta) \Pr(s_t \mid Y_{t-1}, \theta) \right] \right\}, \tag{13}$$

[6]This is true for inference within each model, but as is well known nearly flat priors are not irrelevant to calculating odds ratios across models. We plan to do more analysis of the the sensitivity to priors of our results on comparing model fit.

where

$$f(Y_t \mid Y_{t-1}, s_t, \theta) = (2\pi)^{-\frac{n}{2}} \left|A_0(s_t)\right| \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n} \left[A'_{0,j}(s_t) y_t y'_t A_{0,j} \right.\right.$$
$$\left.\left. -2A'_{+,j}(s_t) x_t y'_t A_{0,j} + A'_{+,j}(s_t) x_t x'_t A_{+,j} \right] \right\},$$

$$\Pr(s_t \mid Y_{t-1}, \theta) = \sum_{s_{t-1}=1}^{h} \left[\Pr(s_t \mid s_{t-1}) \Pr(s_{t-1} \mid Y_{t-1}, \theta) \right].$$

The probability $\Pr(s_{t-1} \mid Y_{t-1}, \theta)$ can be updated recursively. We begin by setting $\Pr(s_0 \mid Y_0, \theta)$ to be the ergodic distribution of the Markov-switching chain. For $t = 1, \ldots, T$, the updating procedure involves the following computation:

$$\Pr(s_t \mid Y_t, \theta) = \frac{f(Y_t \mid Y_{t-1}, s_t, \theta) \Pr(s_t \mid Y_{t-1}, \theta)}{\sum_{s_t=1}^{h} \left[f(Y_t \mid Y_{t-1}, s_t, \theta) \Pr(s_t \mid Y_{t-1}, \theta) \right]}. \tag{14}$$

Multiplying the likelihood function (13) by the prior specified in Section A gives the posterior function $\pi(\theta \mid Y_T)$. When the number of model parameters is small, one could obtain the posterior estimate of θ by simply finding the value of θ that maximizes $\pi(\theta \mid Y_T)$ as in Sims 2001b. But for a fair-size system of simultaneous equations, there can be easily over hundreds of structural parameters and there is no straight maximization routine that can reliably handle such a case. We suggest to use the expectation-maximizing (EM) algorithm proposed by Chib 1996. To keep the number of parameters small and manageable at each maximization step in this EM algorithm, we adopt a Gibbs sampling idea to sample from conditional posterior distributions in addition to sampling unobserved states across time.

Appendix C. Conditional Posterior Distributions

Obtaining the posterior results directly from $\pi(\theta|Y_T)$ proves impossible. One can, however, use a Gibbs sampler to obtain the joint distribution $\pi(\theta, S_T \mid Y_T)$ where $S_t = \{s_1, \ldots, s_t\}$ for $t = 1, \ldots, T$. Denote

$$S^t = \{s_t, \ldots, s_T\};$$
$$Y^t = \{y_t, \ldots, y_T\}.$$

The Gibbs sampler we propose here involves sampling alternatively from the following conditional posterior distributions:

$$\pi(S_T \mid Y_T, \lambda, g, b, p),$$
$$\pi(\lambda \mid Y_T, S_T, g, b, p),$$
$$\pi(g \mid Y_T, S_T, \lambda, b, p),$$
$$\pi(b \mid Y_T, S_T, \lambda, g, p),$$
$$\pi(p \mid Y_T, S_T, \lambda, g, b).$$

Paths of S_T can be simulated recursively backward from the first conditional posterior distribution $\pi(S_T \mid Y_T, \lambda, g, b, p)$ or $\pi(S_T \mid Y_T, \theta)$. To see how this recursion can be done, note that

$$\Pr(S_T \mid Y_T, \theta) = \Pr(s_T \mid Y_T, \theta) \cdots \Pr(s_t \mid Y_T, S^{t+1}, \theta) \cdots \Pr(s_1 \mid Y_T, S^2, \theta); \tag{15}$$

and

$$\begin{aligned}\Pr(s_t \mid Y_T, S^{t+1}, \theta) &\propto \Pr(s_t \mid Y_t, \theta) \Pr(Y^{t+1}, S^{t+1} \mid Y_t, s_t, \theta) \\ &\propto \Pr(s_t \mid Y_t, \theta) \Pr(s_{t+1} \mid s_t, \theta) \Pr(Y^{t+1}, S^{t+2} \mid Y_t, s_t, s_{t+1}, \theta) \\ &\propto \Pr(s_t \mid Y_t, \theta) \Pr(s_{t+1} \mid s_t, \theta),\end{aligned} \tag{16}$$

because $\Pr(Y^{t+1}, S^{t+2} \mid Y_t, s_t, s_{t+1}, \theta)$ is independent of s_t when s_{t+1} is given. Relationship (16) implies that

$$\Pr(s_t \mid Y_T, S^{t+1}, \theta) = \frac{\Pr(s_t \mid Y_t, \theta) \Pr(s_{t+1} \mid s_t, \theta)}{\sum_{s_t=1}^{h} \left[\Pr(s_t \mid Y_t, \theta) \Pr(s_{t+1} \mid s_t, \theta)\right]}. \tag{17}$$

Backward recursion begins by drawing s_T from $\Pr(s_T | Y_T, \theta)$ according to (14) and drawing s_t recursively given the path S^{t+1} according to (17). According to (15), the draws of S_T this way come from $\Pr(S_T | Y_T, \theta)$.

To derive other conditional posterior distributions, we introduce the following notation:

$$\underset{T_k \times n}{Y_k} = \begin{bmatrix} y'_{t_1} \\ \vdots \\ y'_{t_q} \\ \vdots \\ y'_{t_{T_k}} \end{bmatrix}, \quad \underset{T_k \times m}{X_k} = \begin{bmatrix} x'_{t_1} \\ \vdots \\ x'_{t_q} \\ \vdots \\ x'_{t_{T_k}} \end{bmatrix}, \quad \text{for } k = 1, \ldots, h,$$

where T_k is the total number of observations in state k and $s_{t_q} = k$ for $q = 1, \ldots, T_k$. Note that $\sum_{k=1}^{h} T_k = T$. Let

$$\underset{nh \times nh}{\widetilde{\Delta}_0^{-1}} = \operatorname{diag}\left[\left\{Y_k' Y_k - 2\overline{S}' X_k' Y_k + \overline{S}' X_k' X_k \overline{S}\right\}_{k=1}^{h}\right],$$

$$\underset{mh \times nh}{\widetilde{\Delta}_{+0j}} = \operatorname{diag}\left[\left\{\lambda_j(k) \left(X_k' Y_k - X_k' X_k \overline{S}\right)\right\}_{k=1}^{h}\right], \quad \widetilde{\Delta}_{0+j} = \widetilde{\Delta}'_{+0j},$$

$$\underset{nh \times nh}{\widetilde{\widetilde{\Delta}}{}_{0j}^{-1}} = \operatorname{diag}\left[\left\{\zeta_j(k) \left(Y_k' Y_k - 2\overline{S}' X_k' Y_k + \overline{S}' X_k' X_k \overline{S}\right)\right\}_{k=1}^{h}\right],$$

$$\underset{mh \times nh}{\widetilde{\widetilde{\Delta}}_{+0j}} = \operatorname{diag}\left[\left\{\zeta_j(k) \left(X_k' Y_k - X_k' X_k \overline{S}\right)\right\}_{k=1}^{h}\right], \quad \widetilde{\widetilde{\Delta}}_{0+j} = \widetilde{\widetilde{\Delta}}{}'_{+0j},$$

$$\underset{mh \times mh}{\widetilde{\widetilde{\Delta}}{}_{+j}^{-1}} = \operatorname{diag}\left[\left\{\zeta_j(k) X_k' X_k\right\}_{k=1}^{h}\right],$$

where $\operatorname{diag}[\cdot]$ represents a diagonal matrix.

For the variances-only model, we have

$$\pi\left(\zeta_j(k) \mid Y_T, S_T, g, b, p\right) = \text{Gamma}\left(T_k/2 + \alpha_\zeta, 1/\left(\zeta_j^*(k)/2 + 1/\beta_\zeta\right)\right), \tag{18}$$

$$\pi\left(g_j \mid Y_T, S_T, \lambda, b, p\right) = \mathcal{N}\left(\tilde{\tilde{g}}_j, \left(V_j' \widetilde{\widetilde{\Delta}}_{+j}^{-1} V_j + \overline{H}_{+j}^{-1}\right)^{-1}\right), \tag{19}$$

$$\begin{aligned}\pi\left(b \mid Y_T, S_T, \lambda, g, p\right) \propto \left(\prod_{k=1}^{h} |A_0(k)|^{T_k}\right) \exp\Bigg\{-\frac{1}{2}\sum_{i=1}^{n}\Bigg[b_i' \left(U_i' \widetilde{\widetilde{\Delta}}_{0i}^{-1} U_i + \overline{H}_{0i}^{-1}\right) b_i \\ -2 g_i' \left(V_i' \widetilde{\widetilde{\Delta}}_{+0i} U_i\right) b_i \Bigg]\Bigg\},\end{aligned} \tag{20}$$

$$\pi\left(p_k \mid Y_T, S_T, \lambda, g, b\right) = \pi(p_k \mid S_T) = \mathcal{D}(\alpha_{1k} + n_{1k}, \ldots, \alpha_{hk} + n_{hk}), \tag{21}$$

where n_{ik} is the total number of one-step transitions from state k to state i for $i, k = 1, \ldots, h$ in the sequence of states S_T drawn from $\pi(S_T \mid Y_T, \theta)$, and

$$\begin{aligned}\zeta_j^*(k) = {} & a_{0,j}(k)' \left[Y_k' Y_k - 2\overline{S}' X_k' X_k + \overline{S}' X_k' X_k \overline{S}\right] a_{0,j}(k) \\ & - 2 d_j(k)' \left[X_k' Y_k - X_k' X_k \overline{S}\right] a_{0,j}(k) + d_j(k)' X_k' X_k d_j(k),\end{aligned}$$

$$\tilde{\tilde{g}}_j = \left(V_j' \widetilde{\widetilde{\Delta}}_{+j}^{-1} V_j + \overline{H}_{+j}^{-1}\right)^{-1} \left(V_j' \widetilde{\widetilde{\Delta}}_{+0j} U_j\right) b_j,$$

$$T = \sum_{i=1}^{h}\sum_{k=1}^{h} n_{ik}.$$

When the λ_t time varying scale factors apply only to D, the conditional posterior density for P is exactly the same as (21) and the conditional posterior densities for g and b are the same as (19) and (20) except that $\widetilde{\widetilde{\Delta}}_{0j}^{-1}$ and $\widetilde{\widetilde{\Delta}}_{+0j}$ be replaced by $\widetilde{\Delta}_0^{-1}$ and $\widetilde{\Delta}_{+0j}$. As for λ, we have

$$\pi\left(\lambda_j(k) \mid Y_T, S_T, g, b, p\right) = \mathcal{N}\left(\tilde{\lambda}_j(k), \tilde{\sigma}_{\lambda,j}^2(k)\right), \tag{22}$$

where

$$\tilde{\lambda}_j(k) = \left(d_j(k)' \left(X_k' Y_k - X_k' X_k \overline{S}\right) a_{0,j}(k)\right) \tilde{\sigma}_{\lambda,j}^2(k),$$

$$\tilde{\sigma}_{\lambda,j}^2(k) = 1/\left(d_j(k)' X_k' X_k d_j(k) + 1/\sigma_\lambda^2\right).$$

Except for (20), one can simulate draws from all other conditional posterior densities. As for the conditional posterior density of b, we use the Gibbs sampling idea of Waggoner and Zha 2000 to sample b_j one at a time conditional on b_i for $i \neq j$ and other parameters. Unlike constant-parameter simultaneous equation models dealt with by Waggoner and Zha 2000, the posterior density of b_j conditional on all other parameters in our case is nonstandard. We thus use the Metropolis algorithm with the following proposal density for the transition from b_j to b_j':

$$J\left(b_j' \mid b_j, Y_T, S_T, \lambda, g, p, b_1, \ldots, b_{j-1}, b_{j+1}, \ldots, b_n\right) = \mathcal{N}(0, \Xi_j), \tag{23}$$

where b_j' is a proposal draw.

To obtain accurate error bands of functions of structural parameters (like the impulse responses and historical decompositions), we need to normalize both the signs of structural parameters and the labels of states. Given state k ($k = 1, \ldots, h$), we normalize A_0 and A_+ according to Waggoner and Zha 2001. To normalize the labels of states, we look at reduced-form residuals of the interest rate and label the states from the smallest residual to the largest residual.

APPENDIX D. POSTERIOR ESTIMATES

Finding the estimate of θ at the peak of the posterior distribution is a difficult task in our simultaneous equation model. As mentioned in Section B, we use the Gibbs sampling idea to implement the EM algorithm. First, we want to find the estimate of b at the peak of $\pi(b|Y_T)$ through the following Monte Carlo integration (E-step):

$$\mathscr{E}\left(b, b^{old}\right) = \frac{1}{Q}\sum_{q=1}^{Q} \log \pi\left(b_j \mid Y_T, S_T^{(q)}, \lambda^{(q)}, g^{(q)}, p^{(q)}\right),$$

where b^{old} is the last iterated value of b, $S_T^{(q)}$, $\lambda^{(q)}$, $g^{(q)}$, and $p^{(q)}$ are simulated from $\pi\left(S_T, \lambda, g, p \mid Y_T, b^{old}\right)$ with the four conditional densities:

$$\pi(S_T \mid Y_T, b^{old}, \lambda, g, p),$$
$$\pi(\lambda \mid Y_T, b^{old}, S_T, g, p),$$
$$\pi(g \mid Y_T, b^{old}, S_T, \lambda, p),$$
$$\pi(p \mid Y_T, b^{old}, S_T, \lambda, g).$$

The posterior estimate of b can be obtained with the following iterations. Given a starting value for b^{old}, we find the value of b that maximizes $\mathscr{E}\left(b, b^{old}\right)$ (M-step). Then we use the maximum value of b as the next value of b^{old}. Continue the iteration until b converges to a fixed point. Denote this posterior estimate by $\hat{b}$.

Next, we find the estimate of λ at the peak of $\pi(\lambda|Y_T, \hat{b})$, denoted by $\hat{\lambda}$, by iterating the process of obtaining the value of λ maximizing the following function:

$$\mathscr{E}\left(\lambda, \lambda^{old}\right) = \frac{1}{Q}\sum_{q=1}^{Q} \log \pi\left(b_j \mid Y_T, S_T^{(q)}, \hat{b}, g^{(q)}, p^{(q)}\right),$$

where $S_T^{(q)}$, $g^{(q)}$, and $p^{(q)}$ are simulated from $\pi\left(S_T, g, p \mid Y_T, \hat{b}, \lambda^{old}\right)$ with the three conditional densities:

$$\pi(S_T \mid Y_T, \hat{b}, \lambda^{old}, g, p),$$
$$\pi(g \mid Y_T, \hat{b}, \lambda^{old}, S_T, p),$$
$$\pi(p \mid Y_T, \hat{b}, \lambda^{old}, S_T, g).$$

Finally, we find the estimate of $\{g, p\}$ at the peak of $\pi(g, p|Y_T, \hat{b}, \hat{\lambda})$, denoted by $\{\hat{g}, \hat{p}\}$, by iterating the process of maximizing the following function over $\{g, p\}$:

$$\mathscr{E}\left(\{g, p\}, \{g^{old}, p^{old}\}\right) = \frac{1}{Q}\sum_{q=1}^{Q} \log \pi\left(\{g, p\} \mid Y_T, \hat{b}, \hat{\lambda}, S_T^{(q)}\right),$$

where $S_T^{(q)}$ are simulated from $\pi\left(S_T \mid Y_T, \hat{b}, \hat{\lambda}, g^{old}, p^{old}\right)$. We denote the posterior estimate of θ by

$$\hat{\theta} = \{\hat{b}, \hat{\lambda}, \hat{g}, \hat{p}\}.$$

APPENDIX E. MARGINAL LIKELIHOOD

To select a model that best fits to the data, we need to estimate the marginal likelihood $\pi(Y_T)$ for each model and then select the one that give the highest value of $\pi(Y_T)$. The methods proposed by Chib 1995 and Chib and Jeliazkov 2001 make it feasible for us to compute the marginal likelihood. Note that

$$\begin{aligned}\pi(Y_T) &= \int f(Y_T \mid \theta)\pi(\theta)\, d\theta \\ &= \frac{f\left(Y_T \mid \hat{\theta}\right)\pi\left(\hat{\theta}\right)}{\pi\left(\hat{\theta} \mid Y_T\right)},\end{aligned} \tag{24}$$

where $\pi\left(\hat{\theta}|Y_T\right)$ can be computed through the following conditional densities:

$$\begin{aligned}\pi\left(\hat{\theta} \mid Y_T\right) = \prod_{i=1}^{n} &\pi\left(\hat{b}_i \mid Y_T, \hat{b}_1, \ldots, \hat{b}_{i-1}\right) \\ &\pi\left(\hat{d} \mid Y_T, \hat{b}\right)\pi\left(\hat{\lambda} \mid Y_T, \hat{b}, \hat{d}\right)\pi\left(\hat{p} \mid Y_T, \hat{b}, \hat{d}, \hat{\lambda}\right).\end{aligned} \tag{25}$$

The method of Chib and Jeliazkov 2001 enables us to estimate the posterior ordinate

$$\pi\left(\hat{b}_i \mid Y_T, \hat{b}_1, \ldots, \hat{b}_{i-1}\right),$$

which requires simulating $\{b_i, b_{i+1}, \ldots, b_n, d, \lambda, p\}$ with the conditional densities:

$$\begin{gathered}\pi\left(b_i \mid Y_T, \hat{b}_1, \ldots, \hat{b}_{i-1}, b_{i+1}, \ldots, b_n, S_T, d, \lambda, p\right), \\ \pi\left(b_{i+1} \mid Y_T, \hat{b}_1, \ldots, \hat{b}_{i-1}, b_i, b_{i+2}, \ldots, b_n, S_T, d, \lambda, p\right), \\ \vdots \\ \pi\left(b_n \mid Y_T, \hat{b}_1, \ldots, \hat{b}_{i-1}, b_i, \ldots, b_{n-1}, S_T, d, \lambda, p\right), \\ \pi\left(S_T \mid Y_T, \hat{b}_1, \ldots, \hat{b}_{i-1}, b_i, \ldots, b_n, d, \lambda, p\right), \\ \pi\left(d \mid Y_T, \hat{b}_1, \ldots, \hat{b}_{i-1}, b_i, \ldots, b_n, S_T, \lambda, p\right), \\ \pi\left(\lambda \mid Y_T, \hat{b}_1, \ldots, \hat{b}_{i-1}, b_i, \ldots, b_n, S_T, d, p\right), \\ \pi\left(p \mid Y_T, \hat{b}_1, \ldots, \hat{b}_{i-1}, b_i, \ldots, b_n, S_T, \lambda, d\right),\end{gathered}$$

and simulating $\{b_{i+1},\ldots,b_n,d,\lambda,p\}$ with the conditional densities:

$$\begin{gathered}
\pi\left(b_{i+1} \mid Y_T,\hat{b}_1,\ldots,\hat{b}_i,b_{i+2},\ldots,b_n,S_T,d,\lambda,p\right),\\
\pi\left(b_{i+2} \mid Y_T,\hat{b}_1,\ldots,\hat{b}_i,b_{i+1},b_{i+3},\ldots,b_n,S_T,d,\lambda,p\right),\\
\vdots\\
\pi\left(b_n \mid Y_T,\hat{b}_1,\ldots,\hat{b}_i,b_{i+1},\ldots,b_{n-1},S_T,d,\lambda,p\right),\\
\pi\left(S_T \mid Y_T,\hat{b}_1,\ldots,\hat{b}_i,b_{i+1},\ldots,b_n,d,\lambda,p\right),\\
\pi\left(d \mid Y_T,\hat{b}_1,\ldots,\hat{b}_i,b_{i+1},\ldots,b_n,S_T,\lambda,p\right),\\
\pi\left(\lambda \mid Y_T,\hat{b}_1,\ldots,\hat{b}_i,b_{i+1},\ldots,b_n,S_T,d,p\right),\\
\pi\left(p \mid Y_T,\hat{b}_1,\ldots,\hat{b}_i,b_{i+1},\ldots,b_n,S_T,\lambda,d\right).
\end{gathered}$$

For other posterior ordinates in (25), we use the method proposed by Chib 1995. Estimation of $\pi\left(\hat{d} \mid Y_T,\hat{b}\right)$ involves simulating $\{d,\lambda,p,S_T\}$ from $\pi\left(d,\lambda,p,S_T \mid Y_T,\hat{b}\right)$; estimation of $\pi\left(\hat{\lambda} \mid Y_T,\hat{b},\hat{d}\right)$ involves simulating $\{\lambda,p,S_T\}$ from $\pi\left(\lambda,p,S_T \mid \hat{b},\hat{d}\right)$; and $\pi\left(\hat{p} \mid Y_T,\hat{b},\hat{d},\hat{\lambda}\right)$ is estimated by simulating λ,S_T from $\pi\left(p,S_T \mid \hat{b},\hat{d},\hat{\lambda}\right)$. All these simulations can be completed with additional reduced Gibbs runs.

The numerator in (24) can be easily computed. The prior ordinate $\pi\left(\hat{\theta}\right)$ is readily available by direct calculation. It is also straightforward to compute the likelihood ordinate $f\left(Y_T \mid \hat{\theta}\right)$ by (13).

Chib and Jeliazkov 2001 also recommend a method to approximate Monte Carlo standard error of the marginal likelihood estimate. The essence of this approximation involves computing the sample variance of a vector of above-mentioned Monte Carlo integrations required to estimate the denominator (25) (Newey and West 1987).

References

CHIB, S. (1995): "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321.

——— (1996): "Calculating Posterior Distributions and Model Estimates in Markov Mixture Models," *Journal of Econometrics*, 75, 79–97.

CHIB, S., AND I. JELIAZKOV (2001): "Marginal Likelihood From the Metropolis-Hastings Output," *Journal of the American Statistical Association*, 96(453), 270–281.

CHRISTIANO, L. J., M. EICHENBAUM, AND C. L. EVANS (2001): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," Discussion paper, Northwestern University.

CLARIDA, R., J. GALÍ, AND M. GERTLER (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, CXV, 147–180.

COGLEY, T., AND T. J. SARGENT (2001): "Evolving US Post-World War II Inflation Dynamics," *NBER Macroeconomics Annual*.

EDGE, R. M. (2000): "Time to Build, Time to Plan, Habit Persistence, and the Liquidity Effect," International Finance Discussion Paper 673, Federal Reserve Board.

HAMILTON, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), 357–384.

HANSON, M. (2001): "Varying Monetary Policy Regimes: A Vector Autoregressive Investigation," Discussion paper, Wesleyan University.

KIM, C.-J., AND C. R. NELSON (1999): *State-Space Models with Regime Switching*. MIT Press, London, England and Cambridge, Massachusetts.

LEEPER, E., AND T. ZHA (2001a): "Modest Policy Interventions," Discussion paper, Indiana University and Federal Reserve Bank of Atlanta, `http://php.indiana.edu/~eleeper/Papers/lz0101Rev.pdf`.

LEEPER, E. M., AND T. ZHA (2001b): "Empirical Analysis of Policy Interventions," Manuscript, Indiana University.

NEWEY, W. K., AND K. K. WEST (1987): "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

SIMS, C. A. (1998): "Stickiness," *Carnegie-rochester Conference Series On Public Policy*, 49(1), 317–356.

——— (1999): "Drift and Breaks in Monetary Policy," Discussion paper, Princeton University, `http://www.princeton.edu/~sims/`, Presented at a plenary session of the July, 1999 meetings of the Econometric Society, Australasian region.

——— (2001a): "Implications of Rational Inattention," Discussion paper, Princeton University, `www.princeton.edu/~sims`.

——— (2001b): "Stability and Instability in US Monetary Policy Behavior," Manuscript, Princeton University.

SIMS, C. A., AND T. ZHA (1998a): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39(4), 949–968.

WAGGONER, D. F., AND T. ZHA (2000): "A Gibbs Sampler for Structural Vector Autoregressions," Federal Reserve Bank of Atlanta Working Paper 2000-3 (March).

——— (2001): "Likelihood Preserving Normalization in Multiple Equation Models," Working Paper, Federal Reserve Bank of Atlanta.