*

WERE THERE REGIME SWITCHES IN US MONETARY POLICY?

CHRISTOPHER A. SIMS AND TAO ZHA*

ABSTRACT. A multivariate regime-switching model for monetary policy is confronted with US data. The best fit allows time variation in disturbance variances only. With coefficients allowed to change the best fit is with change only in the monetary policy rule and there are three estimated regimes corresponding roughly to periods when most observers believe that monetary policy actually differed. But the differences among regimes are not large enough to account for the rise, then decline, in inflation of the 70's and 80's. Our estimates imply monetary targeting was central in the early 80's, but also important sporadically in the 70's. (*JEL* E52, E47, C53)

It is widely thought that US monetary policy changed a great deal, and for the better, between the 1970's and the 1980's. Richard Clarida, Jordi Gali and Mark Gertler (2000) (CGG) and Thomas A. Lubik and Frank Schorfheide (2004) find that the policy rule apparently followed in the 70's was one that, when embedded in a stochastic general equilibrium models, would imply non-uniqueness of the equilibrium and hence vulnerability of the economy to "sunspot" fluctuations of arbitrarily large size. Their estimated policy rule for the later period, on the other hand, implied no such indeterminacy. These results apparently provide an explanation of the volatile and rising inflation of the 70's and of its subsequent decline.

There are other interpretations of the evidence, however. Giorgio Primiceri (2003) and Thomas J. Sargent, Noah Williams and Tao Zha (in press) estimate models that find only modest changes in policy in the past four decades. Ben Bernanke and Ilian Mihov (1998), Eric M. Leeper and Tao Zha (2003), and James H. Stock and Mark W. Watson (2003) perform several econometric tests and do not find strong evidence against stability of coefficients. An earlier version of this paper (entitled "Macroeconomic Switching") and subsequent studies (Fabio Canova and Luca Gambetti, 2004; Chang-Jin Kim and Charles R. Nelson, 2004; Timothy Cogley and Thomas J. Sargent, 2005; Giorgio Primiceri, 2005) show little evidence in favor of the view that the monetary policy rule has changed drastically.

This paper follows the structural VAR literature in making explicit identifying assumptions to isolate estimates of monetary policy behavior and its effects on the economy, while keeping the model free of the many additional restrictive assumptions needed to give every parameter and equation a behavioral interpretation or to allow structural interpretation of a single-equation model. We use a model that allows explicitly for changes in policy regime, including as special cases both short-lived oscillating policy changes and unidirectional, persistent shifts toward improved policy. We compare versions of the model with Bayesian posterior odds ratios, a method that automatically penalizes models with unneeded free parameters.

Our most important empirical finding is that the version of our model that fits best is one that shows no change at all in coefficients either of the policy rule or of the private sector block of the model. What changes across "regimes" is only the variances of structural disturbances. That is, this version of the model explains differences in the behavior of the economy between periods as reflecting variation in the sources of economic disturbances, not as variation in the dynamics of the effects of a given disturbance on the economy. The Volcker reserves-targeting period emerges as a period of high variance in disturbances of the policy rule. This finding lends empirical support to the common practice in the literature of combining the samples before and after reserve-targeting period to estimate the model, as long as heteroskedasticity is properly taken into account.

We consider also models in which parameters do change. We have looked at models where all parameters in all equations can change, where only non-monetary-policy coefficients change, and where only monetary-policy coefficients can change. In these cases we allow structural variances to shift size at the same time coefficients change, and we have also tried models in which the times of coefficient changes are stochastically independent of the times of variance changes. We have allowed the number of regimes to vary, including the case of a single regime, and we have considered specifications in which regime change is constrained to be monotonic, so that old regimes are constrained never to recur. None of these models fit nearly as well as the best-fitting model in which only residual variances change across regimes. Particularly ill-fitting are the models with a single regime and the model that constrains regime changes to be monotonic.

The best fitting model among those that do allow coefficients to change is one that constrains the changes to occur only in the monetary policy equation, while coefficients in the

other equations remain constant. Like Cogley and Sargent (2005) and Primiceri (2005), we find that the point estimates of the changes are not trivial, even though the data leave their magnitudes uncertain. The model finds the best fit with four regimes. One occurs in only a few brief spans of months, one of which is September-October 2001, and has very high residual variance in money demand. Another corresponds to the Volcker reserve-targeting period and shows clearly the targeting of monetary aggregates, rather than interest rates, in that regime. Another regime has been in place through nearly all of the years of the Greenspan Fed chairmanship — but also was in place through most of the 60's. A fourth regime occurred in several multi-year episodes in the late 60's and early 70's. Though it does not show as strong a monetary-aggregate-targeting flavor as the Volcker regime, it does tend much more strongly in that direction than the "Greenspan" regime. We call this fourth regime the "Burns" regime, even though the "Greenspan" regime was in place though approximately the same proportion of the Burns chairmanship as was the "Burns" regime. (For most of this paper we drop the quotes on the regime names, hoping the reader can bear in mind that the correspondence of the regimes to chairmanship terms is rough.)

We display counterfactual simulations of history with alternate monetary policy regimes. If we simulate history with the estimated time series of shocks, but the coefficients of the policy rule set at the estimated Greenspan policy throughout the period 1961-1987, the rise and fall of inflation follows the historical path extremely closely. This is not because the model is incapable of showing an effect of monetary policy. If we instead use a policy rule that uses the Greenspan coefficients, except that it doubles the coefficients on inflation, the counterfactual historical simulation shows much lower inflation throughout the 70's and early 80's — at the cost of considerably lower output growth through that period. A similar lower inflation path emerges if we fix the policy rule at the point estimate for the Volcker reserve-targeting regime.

Although the estimated differences in policy behavior and their effects on the economy in this four-state model are substantively interesting and consistent with the results from the recent learning literature (Primiceri, 2005; Sargent et al., in press), they are not as drastic as what is implied by the sunspot-equilibrium model. In particular, for all of the three main regimes our estimates imply that with high probability monetary policy responses to inflation were strong enough to guarantee a determinate equilibrium price level.

There are a number of likely explanations for the contrast between our finding here and the findings in some other empirical papers. Perhaps the most important is that rather than aiming at finding some model we can interpret that is not rejected by the data, we aim at fully characterizing the uncertainty about our results. When we run our counterfactual historical simulations by drawing from the posterior distribution of the coefficients of the policy rule instead of fixing the coefficients at particular values, we can see that the shape of uncertainty about these policy rules differs more than do their most likely values. When we simulate history with the Greenspan, Burns and Volcker rule *distributions*, the median paths for inflation and output show visible differences, with the Volcker and Greenspan median paths similar and lower than the Burns median path. The Volcker and Greenspan distributions show a risk of deflation, while the Burns distribution does not, and the Volcker and Greenspan paths show a risk of periods of output growth below -5 per cent at an annual rate, while the Burns path does not. The output growth rate along the median Burns path is slightly above the historical growth rate, while it is notably below ($\frac{1}{2}$ to 1 per cent at annual rate) the historical rate along the Greenspan and Volcker medians. The Burns distribution shows a risk of inflation not coming down at all in the 80's, while neither the Volcker nor the Greenspan path shows such a risk. In other words, even though the data are best explained by a model with no change at all in policy rule coefficients, if one looks for changes, and one is willing to consider policy rules that are unlikely but not impossible, one can tell a story consistent with the view that the Burns policy, had it persisted (instead of ending around 1977, as the model estimates it did), would have failed to end inflation.

There are also substantive differences between our model and the rest of the literature that may contribute to our finding that there is little evidence of policy change. Of particular note is the fact that, unlike much previous work, which fits a "Taylor rule" to the whole period, we include a monetary aggregate in our policy reaction function. The Federal Reserve is by law required to provide the target paths for various monetary aggregates, and during the 70's the behavior of these aggregates was central to discussions of monetary policy. We show that constraining the monetary aggregate not to appear in our monetary policy equation greatly worsens the model's fit to the historical data, and we argue that it is likely that excluding the aggregate from the equation was a source of bias in earlier work. However, while excluding money might have led to a spurious finding of a violation of the "Taylor

principle", including money in our framework improves the relative fit of models that allow variation in the policy rule.

We think our results have implications for future research on theoretical models with more detailed behavioral structure:

- The Taylor rule formalism, valuable as it may be as a way to characterize policy in the last 20 years, can be seriously misleading if we try to use it to interpret other historical periods, where monetary aggregate growth was an important factor in the thinking of policy-makers.

- It is time to abandon the idea that policy change is best modeled as a once-and-for-all, non-stochastic regime switch. Policy changes, if they have occurred, have not been monotonic, and they have been difficult to detect. Both the rational public in our models and econometricians must treat the changes in policy probabilistically, with a model of how and when the policy shifts occur and with recognition of the uncertainty about their nature and timing.

I. The Debate over Monetary Policy Change

The literature in this area is large enough that we will not try to discuss papers in it one by one, but we lay out what seem to us a few of the most important reasons why our results differ from much of the previous empirical work in the area.

(i) As we pointed out above, our specification includes a monetary aggregate in the reaction function. Most of the previous literature does not. We think this is a possibly important source of bias in estimates of the reaction function.

(ii) Much of the previous literature either makes no allowance for heteroskedasticity or allows only implausibly restricted forms of heteroskedasticity. Particularly common have been specifications in which there is a single change in residual variance in the sample, and specifications that generate "robust standard errors" by allowing for heteroskedasticity that is a function of right-hand side variables. It is clear to the eye, and apparent in our estimation results, that residual variances in the reaction function rose sharply at the end of 1979, then dropped back a few years later. A single shift in variance cannot capture this fact. And the persistent shifts in variances that we find could not be well modeled as functions of right-hand-side variables. As

we have already noted, failing to allow properly for heteroskedasticity can strongly bias statistical tests in favor of finding significant shifts in coefficients. This is apparent from the contrast between the results of Cogley and Sargent (2001) and the later version Cogley and Sargent (2005) that does allow for a fairly general form of heteroskedasticity.

(iii) Identification in these models is fragile. This is particularly true for the forward-looking Taylor rule specification of CGG, for two reasons.

One is that estimating this single equation is based on claiming that a list of instrumental variables is available that can be used to control for the endogeneity of expected future inflation and and output. But these instruments are available only because of a claim that we know a priori that they do not enter directly into the reaction function — they can affect monetary policy only through their effects on expected future variables. We find it inherently implausible that, for example, the monetary authority reacts to an expected future 3 per cent inflation rate in exactly the same way, whether the recent past level of inflation has been 1.5 per cent or 6 per cent.

The other problem with this specification is that the Fisher relation is always lurking in the background. The Fisher relation connects current nominal rates to expected future inflation rates and to real interest rates, which are in turn plausibly determined by expected output growth rates. So one might easily find an equation that had the form of the forward-looking Taylor rule, satisfied the identifying restrictions, but was something other than a policy reaction function.

Multivariate models allow a check on the identifying assumptions via examination of the impulse responses to monetary policy shocks. Single equation approaches obviously do not. It seems to us that empirical work that has been based on multivariate models and has included checks for plausibility of responses to monetary policy shocks has tended to find less evidence of changing monetary policy.

(iv) It is interesting to consider changes in monetary policy and to connect estimated changes to historical events. Indeed, we do some of that in this paper, with a model we do not think is our best. As a result, abstracts, introductions and conclusions often seem to support the idea that there have been changes in monetary policy

even when a look at plotted confidence or probability bands around time paths of coefficients or functions of them can be seen to include constant paths. So in some cases there is more contrast between the abstracts of papers in the literature and our abstract than there is in the detailed results.

II. Class of Models

The general framework is described by nonlinear stochastic dynamic simultaneous equations of the form:

$$(1) \qquad y_t' A_0(s_t) = x_t' A_+(s_t) + \epsilon_t', \quad t = 1, \ldots, T,$$

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

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.

Denote the longest lag length in the system of equations (1) by ν. The vector of right-hand variables, x_t, is ordered from the n endogenous variables for the first lag down to the n variables for the last (ν^{th}) lag with the last element of x_t being the constant term.

For $t = 1, \ldots, T$, denote

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

We treat as given the initial lagged values of endogenous variables $Y_0 = \{y_{1-\nu}, \ldots, y_0\}$. Structural disturbances are assumed to have the distribution:

$$\pi(\epsilon_t \mid Y_{t-1}) = \mathcal{N}\left(\underset{n \times 1}{\mathbf{0}}, \mathbf{I}_n\right),$$

where $\mathcal{N}(a, b)$ refers to the normal pdf with mean a and covariance matrix b and $\mathbf{I}_n$ is an $n \times n$ identity matrix. Following James D. Hamilton (1989) and Siddhartha Chib (1996), we impose no restrictions on the transition matrix $P = [p_{ik}]$.

The reduced-form system of equations implied by (1) is:

$$(3) \qquad y_t' = x_t' B(s_t) + u_t'(s_t), \quad t = 1, \ldots, T;$$

where

$$B(s_t) = A_+(s_t)A_0^{-1}(s_t), \tag{4}$$

$$u_t(s_t) = A_0'^{-1}(s_t)\,\epsilon_t, \tag{5}$$

$$E(u_t(s_t)u_t(s_t)') = (A_0(s_t)A_0'(s_t))^{-1}. \tag{6}$$

In the reduced form (4)-(6), $B(s_t)$ and $u_t(s_t)$ involve the structural parameters and shocks across equations, making it impossible to distinguish regime shifts from one structural equation to another. In contrast, the structural form (1) allows one to identify each structural equation, such as the policy rule, for regime switches.

If we let all parameters vary across states, it is relatively straightforward to apply the existing methods of Chib (1996) and Sims and Zha (1998) to the model estimation because $A_0(s_t)$ and $A_+(s_t)$ 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 and 6 endogenous variables, for example, the number of parameters in $A_+(s_t)$ is of order 468 for each state. Given the post-war macroeconomic data, however, 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 restricting the degree of time variation in the model's parameters.[1]

We rewrite A_+ as

$$\underset{m\times n}{A_+(s_t)} = \underset{m\times n}{D(s_t)} + \underset{m\times n}{\overline{S}}\,\underset{n\times n}{A_0(s_t)}. \tag{7}$$

where

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

If we place a prior distribution on $D(s_t)$ that has mean zero, our prior is centered on the same reduced-form random walk model that is the prior mean in existing Bayesian VAR models (Sims and Zha, 1998). As can be seen from (4)-(7), this form of prior implies that smaller A_0^{-1} values, and thus smaller reduced form residual variances, are associated with tighter concentration of the prior about the random walk form of the reduced form. On

the other hand, small values of D are also associated with tighter concentration of the prior about the random walk reduced form, without any corresponding effect on reduced form residual variances.

Note that this setup centers the prior on models in which the moving average representation[2] has the form

$$y_t' = \sum_{s=0}^{\infty} \epsilon_{t-s}' A_0^{-1} \ .$$

This ties our beliefs about lagged effects of structural innovation i on variable j to our beliefs about contemporaneous effects of innovation i on variable j. Any prior that centers on a random walk reduced form while leaving beliefs about reduced form residual covariances independent of beliefs about reduced form coefficients, will have the same effect. For example the standard "Minnesota prior" on the reduced form, combined with any identification scheme based on restrictions on contemporaneous coefficients, will center on MAR's of this form. If one thinks of the model as a discrete approximation to an underlying continuous-time system, this type of prior is reasonable. It is implausible that the effects of structural innovations show sharp discontinuities across lags.

We consider the following three cases of restricted time variation for $A_0(s_t)$ and $D(s_t)$:

$$(8) \qquad a_{0,j}(s_t), d_{ij,\ell}(s_t), c_j(s_t) = \begin{cases} \bar{a}_{0,j}, \bar{d}_{ij,\ell}, \bar{c}_j & \text{Case I} \\ \bar{a}_{0,j}\xi_j(s_t), \bar{d}_{ij,\ell}\xi_j(s_t), \bar{c}_j\xi_j(s_t) & \text{Case II} \\ a_{0,j}(s_t), \bar{d}_{ij,\ell}\lambda_{ij}(s_t), c_j(s_t) & \text{Case III} \end{cases} ,$$

where $\xi_j(s_t)$ is a scale factor for the jth structural equation, $a_{0,j}(s_t)$ is the jth column of $A_0(s_t)$, $d_j(s_t)$ is the jth column of $D(s_t)$, $d_{ij,\ell}(s_t)$ is the element of $d_j(s_t)$ for the ith variable at the ℓth lag, the last element of $d_j(s_t)$, $c_j(s_t)$, is the constant term for equation j. The parameter $\lambda_{ij}(s_t)$ changes with variables but does not vary across lags. This allows long run responses to vary over time, while constraining the dynamic form of the responses to vary only through λ_{ii}, which can be though of as indexing the degree of inertia in the variable interpreted as the "left-hand side". Of course in this simultaneous equations setup, there may not be a variable that is uniquely appropriate as "left-hand side" in equation i. The specification insures, though, that whichever variable we think of as on the left hand side, the time variation in dynamics is one-dimensional, in that it affects all "right-hand side" variables

in the same way. The bar symbol over $a_{0,j}$, $d_{ij,\ell}$, and c_j means that these parameters are state-independent (i.e., constant across time).

Case I is a constant-coefficient structural equation. Case II is an equation with time-varying disturbance variances only. Case III is an equation with time-varying coefficients, as well as time varying disturbance variances.

We have considered models with Case II specifications for all equations, with Case II for the policy equation and Case III for all others, with Case III for the policy equation and Case II for all others, and with Case III for all equations. That is, we have examined models with time variation in coefficients in all equations, with time variation in coefficients in policy or private sector equations only, and with no time variation in coefficients. In all of these cases we allow time variation in structural disturbance variances of all equations. The model with time variation in coefficients in all equations might be expected to fit best if there were policy regime changes and the nonlinear effects of these changes on private sector dynamics, via changes in private sector forecasting behavior, were important. That this is possible was the main point of Robert E. Lucas (1972).

However, as Sims (1987) has explained at more length, once we recognize that changes in policy must in principle themselves be modeled as stochastic, Lucas's argument can be seen as a claim that a certain sort of nonlinearity is important. Even if the public believes that policy is time-varying and tries to adjust its expectation-formation accordingly, its behavior could be well approximated as linear and non-time-varying. As with any use of a linear approximation, it is an empirical matter whether the linear approximation is adequate for a particular sample or counterfactual analysis.[3]

We consider the model with Case III for all equations because we are interested in whether it fits better than the other models, as would be true if policy had changed within the sample and Lucas-critique nonlinearities were important. We consider the other combinations because it is possible that coefficients in the policy have not changed enough for the changes to emerge clearly from the data, or enough to generate detectable corresponding changes in private sector behavior.

III. Data, Identification, and Model Fit

We use monthly US data from 1959:1–2003:3. Each model has 13 lags and includes the constant term and 6 commonly-used endogenous variables: a commodity price index (Pcom), M2 divisia (M), the federal funds rate (R), interpolated monthly real GDP (y), the core personal consumption expenditure (PCE) 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.[4]

The identification of monetary policy, following Leeper and Zha (2003), 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.[5]

In addition to the exact zero restrictions shown in Table 1, we introduce stochastic prior information favoring a negative contemporaneous response of money demand to the interest rate and a positive contemporaneous response of the interest rate to money (see Appendix A). More precisely, we use a prior that makes the coefficients on R and M in the money demand column of A_0 positively correlated and in the monetary policy column of A_0 negatively correlated. This liquidity effect prior has little influence on the correlation of posterior estimates of the coefficients in the policy and the money demand equations, but it makes point estimates of coefficients and impulse responses more stable across different sample periods. The instability we eliminate here arises from the difficulty of separating money demand and supply in some sub-periods, and is associated with imprecise estimates in both equations for this reason. Since a finding of change in monetary policy across periods requires some precision in the estimates of policy rule coefficients in those periods, the liquidity-effect priors are as likely to strengthen as to weaken evidence for changes in the policy rule. We take up this issue again in discussion of the results, below.

We model and compare the five specifications:

Constant: a constant-parameter BVAR (i.e., all equations are Case I);
Variances Only: all equations are Case II;

Monetary Policy: all equations except the monetary policy rule are Case II, while the policy rule is Case III;

Private Sector: equations in the private sector are Case III and monetary policy is Case II;

All Change: all equations are Case III.

There are two major factors that make the estimation and inference of our models a difficult task. One factor is simultaneous relationships in the structural coefficient matrix $A_0(s_t)$. The other factor is the types of restricted time variations specified in (8). Without these elements, the shape of the posterior density would be much more regular and more straightforward Gibbs sampling methods would apply. Appendix A outlines the methods and briefly discusses both analytical and computational difficulties.

The first set of results to consider is measures of model fit, with the comparison based on posterior marginal data densities. The results are displayed in Table 2. For the models with larger numbers of free parameters the Markov Chain Monte Carlo (MCMC) sample averages that are the basis of the numbers in the table behave erratically, and we display "$*$" for these cases rather than a specific number. Though the estimated marginal data densities (MDD's) for these cases are erratic, they remain far below the levels of the MDD's shown in the same column above them. In other words, though displaying a single number for their MDD values might indicate misleading precision, it is clear that the MDD's for these cases are very much lower than those of the cases for which we do display numbers.[6]

Note that this is a log-likelihood scale, so that differences of 1 or 2 in absolute value mean little, while differences of 10 or more imply extreme odds ratios in favor of the higher-marginal-data-density model. For the upper rows in the table the Monte Carlo (MC) error in these numbers (based on two million MCMC draws) is from ± 2 to ± 4. For the lower rows in each column the error is larger (from ± 3 to ± 5). These estimates of MC error are conservative, based on our own experience with multiple starting points for the chain. Conventional measures of accuracy based on serial covariances of the draws, for example, would suggest much smaller error bands. When the whole private sector, or the whole model, is allowed to change according to Case III, the marginal data density is distinctly lower than that of the best models for a given row of the table and for those versions of the model for which we could obtain convergence. The best fit is for the 9-state variances-only model,

though any of the 7 through 10 state versions of that model have similar fit. The marginal data density for these variances-only models are higher by at least 50 on a log scale than that for any other model. The best of the models allowing time variation in coefficients is the monetary policy model with 4 states, whose marginal data density is higher by at least 50 than that of any other model that allows change in coefficients.[7]

IV. Best-Fit Model

There are a number of best-fit models, all of them variances-only models with from 7 to 10 states. Since the results from these models are quite similar, we report the results from only the 9-state variances-only model. The transition matrix for the 9 states is shown in Table 3. The states appear to behave similarly, and they have a fairly evenly spread set of steady-state probabilities, ranging from .078 to .19.

The first state is used as a benchmark with its variances being normalized to 1. As can be seen from Figure 1, this state prevails in most of the Greenspan regime and includes several years in the 1960s. The variances in other states do not simply scale up and down across all structural equations. Some states affect a group of structural shocks jointly, as can be seen from Table 4. The 9th state prevails in the Volcker reserve-targeting period, and primarily inflates the variance of the policy shock (Figure 1 and Table 4.) The 8th state inflates the variances of several private-sector equations, and it prevails only for the two months of September and October, 2001. This is clearly a "9/11" state. The other states exist sporadically over the 70's as well as over the period from 1983 to 1987 and some years in the 60's. Among these states, the shock variances change irregularly from state to state. For the 70's, short-lived states with changing shock variances reflect several economic disruptions (e.g., two big oil shocks) and the ambivalent way monetary policy was conducted in response to those disturbances.

For this variances-only model, the structural parameters and impulse responses vary across states only up to scales. Table 5 reports the estimate of contemporaneous coefficient matrix for the 1st state. As can be seen from the "M Policy" column, the policy rule shows a much larger contemporaneous coefficient on R than on M, implying the Federal Reserve pays much more attention within the month to the interest rate than the money stock.

Estimates of the model's dynamic responses are very similar to those produced by previous identified VAR models, so we will not present a full set of impulse responses. The results are as sensible as for previous models, yet we have a more accurate picture of uncertainty because of its stochastically evolving shock variances. The responses to a monetary policy shock for the 1st state, together with error bands, are shown in Figure 2.[8] Note that, though commodity prices and the money stock decline following a shock that tightens monetary policy, the point estimates show P declining only after a delay of several years, and this decline is small and uncertain.

Table 6 reports artificial long run responses of the policy rate to other macro variables, as often presented in the literature. By "artificial" we mean that these are neither an equilibrium outcome nor multivariate impulse responses, but are calculated from the policy reaction function *alone*, asking what would be the permanent response in R to a permanent increase in the level or rate of change of the variable in question, if all other variables remained constant. The long run response to the level of the variable is calculated as $\sum_{\ell=0}^{\nu} \alpha_\ell / \sum_{\ell=0}^{\nu} \delta_\ell$, where α_ℓ is the coefficient on the ℓth lag of the "right-hand-side" variable and δ_ℓ is the coefficient on the ℓth lag of the "left-hand-side" variable in the policy rule. The long run response to the change of the variable is calculated as $\sum_{\ell=0}^{\nu} \sum_{i=0}^{\ell} \alpha_i / \sum_{\ell=0}^{\nu} \delta_\ell$. In Table 6, the differenced (log) variables such as Δy and ΔP are annualized to match the annual rate of interest R. Absence of sunspots in the price level will be associated with the sum of these long run responses to nominal variables (here $\Delta PCom$, ΔM, and ΔP) exceeding 1. For this model the sum is 1.76, well above one, though the error bands on individual coefficient leave room for some uncertainty.

V. Policy Regime Switches

In this section, we present the key results from the 4-state model with time-varying coefficients in the policy rule. There are two reasons why this model may be of interest, despite the fact that it is dominated in fit by the model with only disturbance variances changing. First, this model's fit is substantially better than all other models that allow change in coefficients (Table 2). Second, the model reflects a prevailing view that the endogenous component of US monetary policy has changed substantially since 1960 and its simulated results capture

some important aspects of conventional wisdom about policy changes from the 70's through the 80's to the 90's.

Figure 3 shows the implied state-probabilities over time produced by this 4-state model. We can see that state 1 has prevailed for most of our full sample period and for the entire period from the late 80's onward. We call this state the "Greenspan" state of policy, but of course one needs to bear in mind that this policy regime was dominant in most of the 60's and in the latter half of the 70's as well. State 2 is the next most common, occurring most frequently from the early 60's through the early 70's (the first oil shock period), though with no sustained periods of prevalence that match those of state 1. We call this the "Burns" regime, even though it matches up with Burns's chairmanship even less well than the "Greenspan" regime matches with Greenspan's. State 3 prevails during the Volcker reserve targeting period and nowhere else except one very brief period around 1970. State 4 occurs only for a few isolated months, including 9/11, and seems clearly to be picking up outliers rather than any systematic change of coefficients.

The estimate of the transition matrix is shown in Table 7. The 4 states behave quite differently. Nearly half of the steady-state probability (0.49) goes to the Greenspan state. For the other half, the probability is 0.25 for the Burns state, 0.143 for the Volcker state, and 0.116 for the fourth state. From Table 7 one can also see that the probability of switching from the Greenspan and Burns states to the Volcker and fourth states is reduced by one half as compared to the probability of switching the other way.

Differences in the contemporaneous coefficient matrix show up across states as well. In Table 8 we can see that the Greenspan regime's contemporaneous coefficient matrix is broadly similar to that estimated for the full sample with the variances-only model (Table 5). In particular, both policy rules show a much larger contemporaneous coefficient on R than on M. On the other hand, we see from Tables 9 and 10 that the Burns and Volcker states both have much larger contemporaneous coefficients on M, with the M coefficient being relatively largest for the Volcker state. These results are consistent with the observation that Burns seemed to pay a lot of attention to money growth in the early 70's and less (more) attention to money growth (the interest rate) in the last few years of his tenure (Burns, 1987; Chappell et al., 2005) and that Greenspan made the interest rate the explicit policy instrument.

The long run policy responses to macro variables show a similar pattern, as reported in Table 11. The Greenspan regime shows slightly stronger point estimates of the responses of the funds rate to money growth and inflation than those implied by the variances-only model (Table 6), but with greater uncertainty because of the smaller effective sample period. For the Volcker and Burns regimes the responses of the federal funds rate are, variable by variable, so ill-determined that we instead present responses of money growth, which seems closer to the short-run policy target in those regimes. We see that the Volcker regime makes money unresponsive to all variables (measured by both point estimates and error bands). The Burns regime shows a disturbingly high responsiveness of money growth to inflation, though the point estimate is still below 1, which is only partially offset by a negative response to the rate of change in commodity prices.

Because the Burns regime looks like the most likely candidate for a potential sunspot incubator, we tried normalizing that regime's reaction function on the interest rate and calculating its long-run response to the sum of the coefficients on all nominal variables — the rate of change in commodity prices, money growth, and inflation. This response is surprisingly well-determined, probably because of collinearity in the sample among the nominal variables.[9] The 68% probability band is (.94,3.50), which makes it very likely that the regime was not a sunspot incubator.

VI. Historical counterfactuals

As a way to quantify the importance of policy change over time, the 4-state time-varying model makes it an internally coherent exercise to calculate what would have happened if regime changes had not occurred, or had occurred when they otherwise didn't, at particular historical dates. We have run quite a few of these experiments, but the main conclusion is that the estimated policy changes do make a noticeable difference, but not a drastic difference. In the following we display examples that seem most relevant to the debate on the effects of monetary policy changes.

VI.1. **Suppressing policy shocks.** The first and simplest of our counterfactual simulations sets the disturbances in the policy equation to zero in the 9-state model. Disturbances and coefficients are otherwise set at high-likelihood values, so that if the policy rule disturbances had been left in place, the simulations would have shown a perfect fit. As can be seen from

Figure 4, the model leaves the time path of inflation almost unchanged. Policy shocks play a crucial role only in attributing the fluctuations of the funds rate in the late 1970's and the early 1980's. The history of inflation is attributed almost entirely to non-policy sources, though of course feeding systematically through a fixed monetary policy rule.

VI.2. **Keeping a fixed Greenspan or Volcker rule in place throughout.** If we run a similar simulation but with the 4-state monetary policy model by placing the estimated Greenspan rule through the pre-Greenspan period 1961-1987, we obtain the results shown in Figure 5. This simulation tracks history almost as well as the previous one. Thus the model attributes the rise and fall in inflation neither to monetary policy shocks nor to changes in in policy regime. In particular, the model reproduces the high peak inflation rates of the early 80's even though the Greenspan reaction function is in place throughout.

With the Burns policy in place throughout this history instead, we obtain the counter-factual history shown in Figure 6. This simulation also reproduces history very closely, matching the rise and the subsequent fall in inflation. This policy keeps inflation slightly lower in the 60's and 70's, but then in the mid-80's lets the inflation level out at a somewhat higher value.

The modest differences across these policies do not mean the model implies that no changes in monetary policy could have prevented a rise in inflation to near double-digit levels. Though the Volcker reaction function is estimated imprecisely because of the short period in which it prevailed, if we repeat our exercise with the point estimate of the Volcker policy function in place, we obtain the results in Figure 7. This policy would have kept money growth much lower, would have kept inflation lower by around two percentage points at its peak, and would have lowered average output growth. Although the output effect may be difficult to see from Figures 5-7, Table 12 shows the substantial implied differences in output growth for the three regime point estimates throughout this entire period.

These results are not reflective simply of the Volcker policy's focus on growth of monetary aggregates. If we simply double the coefficients on inflation in the Greenspan monetary policy rule, while again leaving disturbances in other equations at historical values and suppressing monetary policy shocks, we arrive at Figure 8. Peak inflation is cut nearly in half, and the inflation rate hovers around zero for much of the 1961-1987 period.

Without any a priori imposed structure on private sector behavior, the model nonetheless shows a type of neutrality result. By the 80's, even though inflation is running 4 or 5 percentage points below the actual historical values, with this "inflation hawk Greenspan" policy output is tracking the historical values almost perfectly. The model thus appears to allow for the public's learning that a new, lower level of inflation prevails. On the other hand, the tighter monetary policy cuts output growth starting in the early 60's, and keeps it well below historical values for most of the 60's, 70's, and 80's. Both of these two policy rules that lower the inflation rate also lower the output growth rate, as can bee seen from Figures 7 and 8.

The counterfactual simulations that imply lower inflation create a marked change in the stochastic process followed by output and inflation. It is therefore quite possible that the output costs of the stronger anti-inflationary policy stance would not have been so persistent as shown in the graphs. Our point is not that stricter anti-inflationary policy would have had output costs as great as shown in these graphs. Our point is only that if the Greenspan rule had been different enough to prevent the rise in inflation in the 60's and 70's, our model would have shown the regime change made a difference. In fact, our best estimate is that the monetary policy regime of the late 80's and 90's was not enough different from the policy actually in place in 60's and 70's to have made any substantial difference to the time path of inflation.

VI.3. **Distributions of policy functions.** Though the policy rules in place before the end of 1979 and after 1982 are estimated to have similar consequences for the rise and fall in inflation, the estimates leave uncertainty about those policies. Point estimates for both regimes show, as we noted above, cumulative responses of the funds rate to inflation that imply a unique price level. Nonetheless the Burns regime point estimates are lower, and the uncertainty about the estimates leaves more probability in the region around a unit response than with the Greenspan regime. As might be expected, the model's simulated time paths respond nonlinearly as the region with less than unit cumulative response of the funds rate to inflation is approached. As a result, if we conduct our counterfactual simulations by drawing from the distribution of policy rule coefficients for the Burns and Greenspan regimes, rather than simply imposing the most likely values, differences between the coefficients become more apparent. In the simulations we describe below the historical shocks are kept on their

historical path, with variances changing with regime according to our estimated posterior distribution, but the policy regime distribution is kept fixed in one regime for all coefficients in the policy equation. This means that the scale of monetary policy shocks, as well as the coefficients in the reaction function, are being drawn from the distribution corresponding to a single regime.

For the Greenspan regime results are shown in Figure 9, where we see that the median simulated path displays substantially lower inflation than what was historically observed. It is important to bear in mind that this is not the actual path for any one policy. This is clear when we look at the median path for interest rates, which is almost uniformly lower than the historical path. If these median paths were actual paths for any given policy, it would be a mystery how the policy could lower inflation without ever raising interest rates. But as can be seen from the graphs for point-estimate policies, policies that lower inflation raise interest rates in some crucial periods, and this is followed by long periods of lower inflation, and hence of lower nominal interest rates. When we display the median path across many policies that imply periods of tighter policy, but imply different timing for the periods of tighter policy, we see a uniformly lower path of interest rates.

Note that these simulated draws from the Greenspan policy distribution imply a substantial risk of deflation in the 80's, as well as a risk of output growth below -5%.

A similar exercise with the Burns regime distribution produces the results in Figure 10. There is little risk of output loss; money growth tends to be higher than the historical path. The risk of deflation is lower, but now there is a substantial risk of no decline at all in inflation in the 80's, consistent with the conventional view about the effects of the Burns policy.

VII. Robustness Analysis

In this section we study a number of other relevant models to check the robustness of our results. The insights from these exercises reinforce the points made in the previous sections.

VII.1. **The economy with policy changes.** We consider an economy with two monetary policy rules estimated in our 4-state policy-only model: one is the rule associated with the Burns regime and the other rule is the Greenspan interest-smoothing policy. This economy consists of the same 6 variables as our actual data and starts with the Burns policy which

lasts for 236 months (corresponding to September 1979 in our sample) and then monetary policy switches, once for all, to the estimated Greenspan policy rule. At the time of the switch in policy rules, the scale of non-policy shocks also changes as in our estimated 4-state model. We simulated 10 samples, each with the same sample length as our actual data and each with initial values set at the actual data from 1959:01 to 1960:01. For each simulated data set, we consider four models: monetary policy models for 2 and 3 states and variances-only models for 2 and 3 states.[10]

In 8 out of the 10 data sets, the estimated transition matrix for the 2-state monetary policy model has one absorbing state, which is of course correct in the simulated data.[11] Thus the method we have used would have been likely to detect a permanent regime shift if that is what had occurred.

Figure 11 shows the cdf, across the 10 Monte Carlo samples, of the posterior probability that there was a change in policy coefficients. In 7 of 10 cases the posterior probability of a change was over .99. In one it was around .2, and in two it was .02 to .03. The log odds ratio corresponding to the most extreme odds against the policy change (i.e., in favor of a variance-change-only model) was 3.78. The log odds ratio in favor of variances-only in our analysis of the historical data is about 60, many times stronger than the most extreme finding in these Monte Carlo simulations.

It is also worth noting that the results showed no tendency to favor spurious variance-change states. The variances-only model with three states had posterior probability less than 10^{-6} in all 10 simulations. The posterior probability on the 3-state model with policy change (which of course is over-parameterized, but contains the true model) reached a maximum of around .04 in one simulation, and otherwise was even smaller than the posterior probability on the 3-state variances-only model.

These experiments give our methods a stiff test. The estimated Greenspan and Burns policy rules that we use leave the qualitative behavior of the data very similar in our counterfactual simulations with point estimators. Yet even with these two similar policy rules, our method is able to detect the switch for a majority of samples.

VII.2. **Other relevant models.**

VII.2.1. *Independent coefficient and variance states.* The results so far assume that coefficients and variances switch at the same time. For the monetary policy model, the potential

problem with this approach is that the number of states for the coefficients on the policy equation must increase with those for the variance state. In a single equation model, Sims (2001) found that making the transitions of variance and coefficient states independent delivered the best fit. In our framework, this can be done by giving special structure to the transition matrix P. If there are two independently evolving state variables, one indexing variances and one indexing equation coefficients, and the transition matrices for the two types of state are Q_1 and Q_2, we get the desired independent evolution by treating each pair of values for the two states as a single state and setting $P = Q_1 \otimes Q_2$.

Estimating a set of models with independent mean and variance states at the same scale of parameterization as our main models would be a major computational task, which we have not undertaken. We have instead calculated maximum log posterior density (LPD) values ("LPD" values rather than log likelihood (LLH) values) for a number of somewhat smaller scale models of this type which we can label 2v, 2v2p, 3v, 3v2p, and 4v. The "nv" models are models with n variance states and no policy coefficient changes. The "nv2p" models are models with n variance states and 2 policy rule coefficient states, evolving independently. Because we have only LPD's, we can't compute posterior odds, but we can (as Sims did in his single-equation paper) compare the models by the Schwarz criterion.[12] The best of the models by this criterion is the 4v model. With the 2v model as base (therefore with the zero value), the Schwarz criteria are:

2v	2v2p	3v	3v2p	4v
0.0	11.1	91.7	78.7	127.9

From this pattern of results it appears that a model with just two coefficient policy regimes is not competitive with variance-only models, even if the variance changes are allowed to evolve independently of the coefficient regimes.

Note that these results may explain why previous researchers (Lubik and Schorfheide, 2004; Clarida et al., 2000, e.g.) who allow only a single change in residual variances find evidence of coefficient change. Those studies are making a comparison like our 2v versus 2v2p comparison in the table, which favors 2v2p. It is only when we allow at least 3 variance states that the addition of a coefficient state ceases to improve fit.

VII.2.2. *Permanent policy shifts.* Our experiments with artificial data suggest that our specification could identify a permanent policy shift if it occurred. Because it is a widespread

view that there was a single permanent shift in US monetary policy around 1979, however, it may nonetheless be of interest to see what emerges if we economize on parameters by imposing on our model the requirement that there is an absorbing state — that is, there is a state that once is entered remains in place for the rest of the sample. This is equivalent to requiring that one column of the transition matrix, which represents the probability of entering each state conditional on being in this state, is a unit vector with a one at the diagonal position.

The fourth column of Table 13 displays the marginal marginal data densities of the monetary policy models with permanent changes on the coefficients of the policy equation. Comparing to the third column of Table 2, we see that the log posterior weight on these models is lower by at least 60 more than the log posterior weights on the models that do not impose the absorbing state restriction.

VII.2.3. *Excluding the monetary aggregate.* In Section V, we have shown the importance of including a monetary aggregate to describe the policy rule under the Burns and Volcker regimes. Here we exclude this variable from the policy reaction function to see if this worsens the fit. The third column of Table 13 reports the measures of fit for a model with 4 states, allowing the monetary policy rule only to change coefficients, and with as usual variances allowed to change with the state in all other equations. The fit is considerably worse than the corresponding cases when money is included (see the third column of Table 2), by about 60 in log odds units.

The fit is also worse when we exclude money from the reaction function in the variances-only model, but the odds ratio is much less extreme. The log odds difference between the 4-state variances-only model and the version of that model with money excluded from the reaction function is 4.46. This implies an odds ratio in favor of the model including money of over 80 in *unlogged* units, but this ratio is much less extreme than the result for the model that allows coefficient variation in the monetary policy rule. This is not surprising, since the most salient difference among the three main estimated policy reaction functions is in the degree to which they give weight to a monetary aggregate. If we shut down this type of difference among policies, the model with coefficient variation in the policy rule is penalized much more than the model that fits a single rule to the whole sample. As we have already pointed out, it seems possible that a model whose prior focused the search for policy

variation in particular economically reasonable directions might be more competitive with the variances-only model. But the results here suggest that such a model, if it is possible at all, is not likely to succeed if it excludes money from the reaction function.

VIII. Conclusion

Monetary policy and its history are complex, and abstract theoretical models that we use to organize thought about them can hide what was really going on. Explorations of data with relatively few preconceptions, like this exploration, may bring out regularities that have been slipping through abstract discussion. In this case, we think this has happened.

Our best-fit model suggests that neither additive disturbances to a linear monetary policy reaction function nor changes in the coefficients of that function have been a primary source of the rise and fall of inflation over our sample period. Instead, stable monetary policy reactions to a changing array of major disturbances generated the historical pattern. Oil price shocks and the Vietnam war and its financing produced disturbances in the 60's and 70's that have not recurred on such a scale since. With such a large role assigned to "private sector shocks", it would be useful to consider a model that allows more detailed interpretation of these shocks. Recent work by Gambetti et al. (2005) is an attempt in this direction.

Even if one gives all the prior weight to the four-state policy model, which assumes the existence of regime changes in monetary policy, our point estimates imply that the impact on the economy of changes in the systematic part of monetary policy were not as big as commonly thought. Nonetheless our estimates do imply that a permanent reserve-targeting policy like that of 1979-82, or a policy that greatly amplified the reaction of interest rates to inflation, could have kept inflation substantially lower, while exacting a cost in lower output growth.

In our estimates that enforce changes in policy rule, the strongest evidence for monetary policy change is that for shifting emphasis on monetary aggregates in the policy reaction function. This accords with the prominent role monetarism played in policy discussions of the 70's. If further research succeeds in finding clear evidence of changes in monetary policy behavior in this period, it will most likely be through focusing attention on the changing impact of monetarism on policy behavior.

Policy actions were difficult to predict, and if there were shifts in the systematic component of policy, they are of a sort that it is difficult for us to track precisely even with hindsight. While our results leave room for those with strong beliefs that monetary policy changed substantially to maintain those beliefs, it is nonetheless clear that whatever changes there were of uncertain timing, not permanent, and not easily understood even today. Models that treat policy changes as permanent, non-stochastic, transparent regime changes are not useful in understanding this history.

Appendix A. Estimation and Inference

A.1. **The Prior.** The identification specified in Table 1 is a special case of standard linear restrictions imposed on A_0 and D as

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

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

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

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 the corresponding free parameters, columns of U_j are orthonormal vectors in the Euclidean space $\mathbb{R}^{nh}$, and columns of V_j are orthonormal vectors in $\mathbb{R}^{mh}$.

The prior distributions for the free parameters b_j and g_j have the following Gaussian forms:

$$\pi(b_j) = \mathcal{N}(\mathbf{0}, \overline{H}_{0j}),$$

$$\pi(g_j) = \mathcal{N}(\mathbf{0}, \overline{H}_{+j}),$$

For all the models studied in this paper, we set H_{0j} and H_{+j} the same way as Sims and Zha 1998 but scale them by the number of states (h) so that the Case I model in (8) coincides with the standard Bayesian VAR with constant parameters. The liquidity effect prior is implemented by adjusting the off-diagonal elements of H_{0j} that correspond to the coefficients of M and R for $j = 2, 3$ such that the correlation for the policy equation (the second equation) is -0.8 and the correlation for the money demand equation (the third

equation) is 0.8. Because we use monthly data, the tightness of the reference prior is set as, in the notation of Sims and Zha 1998, $\lambda_0 = 0.6, \lambda_1 = 0.1, \lambda_2 = 1.0, \lambda_3 = 1.2, \lambda_4 = 0.1, \mu_5 = 5.0$, and $\mu_6 = 5.0$ (see Robertson and Tallman 2001).

The prior distribution for $\xi_j(k)$ is taken as $\pi(\zeta_j(k)) = \Gamma(\alpha_\zeta, \beta_\zeta)$ for $k \in \{1, \ldots, h\}$, where $\zeta_j(k) \equiv \xi_j^2(k)$ and $\Gamma(\cdot)$ denotes the standard gamma pdf with β_ζ being a scale factor (not an inverse scale factor as in the notation of some textbooks). The prior pdf for $\lambda_{ij}(k)$ is $\mathcal{N}(0, \sigma_\lambda^2)$ for $k \in \{1, \ldots, h\}$.

The prior of the transition matrix P takes a Dirichlet form as suggested by Chib 1996. For the kth column of P, p_k, the prior density is

$$\pi(p_k) = \pi(p_{1k}, \ldots, p_{hk}) = \mathcal{D}(\alpha_{1k}, \ldots, \alpha_{hk}) \propto p_{1k}^{\alpha_{1k}-1} \cdots p_{hk}^{\alpha_{hk}-1},$$

where $\alpha_{ik} > 0$ for $i = 1, \ldots, h$.

The hyperparameters α_ζ, β_ζ, and σ_λ are newly introduced and have no reference values in the literature. We set $\alpha_\zeta = \beta_\zeta = 1$ and $\sigma_\lambda = 50$ as the benchmark and then perform a sensitivity check by varying these values. The prior setting $\sigma_\lambda = 50$ is reasonable because the posterior estimate of $\lambda_{ij}(k)$ can be as large as 40 or 50 even with a much smaller value of σ_λ.[13]

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 $\upsilon_{kk} = 0.95$ and $\upsilon_{ik} = 0.05/(h-1)$ for $i \neq k$. Note that $\sum_{i=1}^{h} \upsilon_{ik} = 1$. In the second step, given υ_{ik} and $\sqrt{\text{Var}(p_{kk})}$ (which is set to 0.025), we solve for α_{kk} through a third polynomial and then for all other elements of the vector α_k through a system of $h-1$ linear equations. This prior expresses the belief that the average duration of each state is about 20 months. We also experienced with different prior values for P, including a very diffuse prior for P by letting υ_{ik} be evenly distributed across i for given k and by letting the prior standard deviation of p_{ik} be much larger than 0.025. The results seem insensitive to these prior values.

A.2. **Posterior Estimate.** We gather different groups of free parameters as follows, with the understanding that we sometimes interchange the use of free parameters and original

(but restricted) parameters.

$$p = \{p_k, \ k = 1, \ldots, h\};$$

$$\gamma = \begin{cases} \zeta = \{\zeta_j(k), \ j = 1, \ldots, n, \ k = 1, \ldots, h\}, & \text{for Case II;} \\ \lambda = \{\lambda_{ij}(k), \ i, j = 1, \ldots, n, \ k = 1, \ldots, h\}, & \text{for Case III;} \end{cases}$$

$$g = \{g_j, \ j = 1, \ldots, n\};$$

$$b = \{b_j, \ j = 1, \ldots, n\};$$

$$\theta = \{p, \gamma, g, b\}.$$

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

From the Bayes rule, the posterior distribution of θ conditional on the data is

$$\pi(\theta \mid Y_T) \propto \pi(\theta)\pi(Y_T \mid \theta),$$

where the prior $\pi(\theta)$ is specified in Section A.1.

In order to avoid very long startup periods for the MCMC sampler, it is important to begin with at least an approximate estimate of the peak of the posterior density $\pi(\theta \mid Y_T)$. Moreover, such an estimate is used as a reference point in normalization to obtain likelihood-based statistical inferences. Because the number of parameters is quite large for our models (over 500), we used an eclectic approach, combining the stochastic expectation-maximizing algorithm with various optimization routines. For some models, the convergence took about 15 hours on an Intel Pentium 4 2.0GHz PC; for others, it took as long as a week.[14]

A.3. **Inference.** Our objective is to obtain the posterior distribution of functions of θ such as impulse responses, forecasts, historical decompositions, and long-run responses of policy. It involves integrating over large dimensions many highly nonlinear functions. We use a Gibbs sampler to obtain the joint distribution $\pi(\theta, S_T \mid Y_T)$ where $S_T = \{s_0, s_1, \ldots, s_T\}$. The Gibbs sampler involves sampling alternatively from the following conditional posterior

distributions:

$$\Pr(S_T \mid Y_T, p, \gamma, g, b),$$

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

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

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

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

It has been shown in the literature that such a Gibbs sampling procedure produces the unique limiting distribution that is the posterior distribution of S_T and θ (e.g., Geweke 1999). The probability density functions of these conditional distributions are quite complicated but can be nonetheless simulated.

A.4. **Normalization.** To obtain accurate posterior distributions of functions of θ (such as long run responses and historical decompositions), we must normalize both the signs of structural equations and the labels of states; otherwise, the posterior distributions will be symmetric with multiple modes, making statistical inferences of interest meaningless. Such normalization is also necessary to achieve efficiency in evaluating the marginal likelihood for model comparison.[15] For both purposes, we normalize the signs of structural equations the same way. Specifically, we use the Waggoner and Zha (2003) normalization rule to determine the column signs of $A_0(k)$ and $A_+(k)$ for any given $k \in \{1, \ldots, h\}$.

Two additional normalizations are (1) scale normalization on $\zeta_j(k)$ and $\lambda_j(k)$ and (2) label normalization on the states. We simulate MCMC posterior draws of θ with $\zeta_j(k) = 1$ and $\lambda_j(k) = \mathbf{1}_{h \times 1}$ for all $j \in \{1, \ldots, n\}$, and $k \in \{1, \ldots, h\}$, where the notation $\mathbf{1}_{h \times 1}$ denotes the $h \times 1$ vector of 1's. For each posterior draw, we label the states so that the posterior probabilities of each state for all $t \in \{1, ..., T\}$ match closest to the posterior estimates of those probabilities.[16]

To estimate the marginal data density $\pi(Y_T)$ for each model, we apply both the modified harmonic mean method (MHM) of Gelfand and Dey (1994) and the method of Chib and Jeliazkov (2001). The MHM method is quite efficient for most models considered in this paper, but it may give unreliable estimates for some models whose posterior distributions

have multiple modes. In such a situation, we also use the Chib and Jeliazkov to check the robustness of the estimate.

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

Table 2. Comprehensive measures of fit

Log marginal data densities

Constant	12,998.20			
	Variances Only	Monetary Policy	Private Sector	All Change
2 states	13,345.71	13,383.36	13,280.74	13308.80
3 states	13,434.25	13,446.13	13,380.77	13426.78
4 states	13,466.86	13,480.18	∗	∗
5 states	13,455.26	13,400.10	∗	∗
6 states	13,510.31	∗	∗	∗
7 states	13,530.71	∗	∗	∗
8 states	13,540.32	∗	∗	∗
9 states	13,544.07	∗	∗	∗
10 states	13,538.03	∗	∗	∗

TABLE 3. Transition matrix for 9-state variances-only model

0.9643	0.0063	0.0117	0.0064	0.0108	0.0057	0.0107	0.0095	0.0049
0.0030	0.9394	0.0047	0.0070	0.0210	0.0062	0.0061	0.0069	0.0112
0.0104	0.0159	0.9455	0.0064	0.0046	0.0063	0.0064	0.0096	0.0057
0.0026	0.0043	0.0042	0.9476	0.0040	0.0058	0.0056	0.0062	0.0051
0.0058	0.0155	0.0044	0.0068	0.9425	0.0185	0.0058	0.0064	0.0057
0.0027	0.0056	0.0058	0.0064	0.0051	0.9406	0.0120	0.0062	0.0050
0.0052	0.0042	0.0081	0.0068	0.0040	0.0057	0.9423	0.0062	0.0053
0.0033	0.0041	0.0069	0.0062	0.0038	0.0056	0.0054	0.9429	0.0049
0.0026	0.0046	0.0087	0.0065	0.0042	0.0056	0.0056	0.0062	0.9522

TABLE 4. Relative shock standard deviations across states for 9-state variances-only model

	Financial	M Policy	M demand	Private y	Private P	Private U
First state	1.00	1.00	1.00	1.00	1.00	1.00
Second state	0.95	1.47	1.03	2.07	1.19	1.69
Third state	1.28	1.65	1.84	1.11	1.12	0.91
Fourth state	2.01	2.65	1.93	1.59	1.29	1.37
Fifth state	1.38	2.95	1.24	1.01	0.96	1.17
Sixth state	2.67	2.99	2.32	2.52	0.95	2.13
Seventh state	2.40	4.43	1.21	1.59	2.58	1.05
Eighth state	2.55	4.49	11.44	4.10	10.48	2.67
Ninth state	1.49	12.57	1.53	1.44	1.48	1.44

TABLE 5. Contemporaneous coefficient matrix for 9-state variances-only model

	Financial	M Policy	M demand	Private y	Private P	Private U
Pcom	70.64	0.00	0.00	0.00	0.00	0.00
M	9.21	−130.24	−669.91	0.00	0.00	0.00
R	−27.30	688.52	−70.10	0.00	0.00	0.00
y	−14.21	0.00	19.85	308.75	−20.77	51.94
P	−5.54	−0.00	216.07	0.00	−1061.30	32.38
U	82.37	0.00	0.00	0.00	0.00	766.38

Table 6. Long run policy responses in 9-state variances-only model

Responses of R to	Posterior peak estimate	.68 probability interval
Δ Pcom	0.21	(0.17, 0.73)
Δ M	0.16	(-0.48, 0.44)
Δ y	0.71	(0.69, 3.36)
Δ P	1.39	(0.45, 2.21)
U	-1.01	(-2.80, -0.42)

Table 7. Transition matrix for 4-state policy-only model

0.9627	0.0460	0.0203	0.0334
0.0214	0.9388	0.0195	0.0174
0.0077	0.0073	0.9414	0.0238
0.0082	0.0079	0.0188	0.9254

Table 8. Contemporaneous coefficient matrix for 1st state in 4-state policy-only model

	Financial	M Policy	M demand	Private y	Private P	Private U
Pcom	68.03	0.00	0.00	0.00	0.00	0.00
M	34.19	-208.60	-559.30	0.00	0.00	0.00
R	-32.62	559.48	-172.64	0.00	-0.00	0.00
y	-4.49	0.00	11.87	272.37	-17.51	51.94
P	8.65	0.00	-54.58	0.00	-1029.19	25.45
U	84.70	0.00	0.00	0.00	0.00	705.57

TABLE 9. Contemporaneous coefficient matrix for 2nd state in 4-state policy-only model

	Financial	M Policy	M demand	Private y	Private P	Private U
Pcom	38.20	0.00	0.00	0.00	0.00	0.00
M	19.20	−221.50	−401.63	0.00	0.00	0.00
R	−18.32	188.29	−123.97	0.00	−0.00	0.00
y	−2.52	0.00	8.52	206.87	−13.72	42.40
P	4.86	0.00	−39.19	0.00	−806.18	20.77
U	47.56	0.00	0.00	0.00	0.00	576.00

TABLE 10. Contemporaneous coefficient matrix for 3rd state in 4-state policy-only model

	Financial	M Policy	M demand	Private y	Private P	Private U
Pcom	50.43	0.00	0.00	0.00	0.00	0.00
M	25.35	−393.51	−241.46	0.00	0.00	0.00
R	−24.18	136.05	−74.53	0.00	−0.00	0.00
y	−3.33	0.00	5.12	235.35	−12.82	41.12
P	6.41	0.00	−23.56	0.00	−753.62	20.15
U	62.78	0.00	0.00	0.00	0.00	558.70

TABLE 11. Long run policy responses in 4-state policy-only model

First state (Greenspan)

Responses of R to	Posterior peak estimate	.68 probability interval
Δ Pcom	0.09	(-0.19, 0.24)
Δ M	0.23	(-0.46, 2.08)
Δ y	0.43	(-1.28, 0.64)
Δ P	1.99	(-0.09, 2.48)
U	-1.29	(-0.91, 0.46)

Second state (Burns)

Responses of Δ M to	Posterior peak estimate	.68 probability interval
Δ Pcom	-0.24	(-0.50, 0.01)
R	0.09	(-0.02, 0.49)
Δ y	0.18	(-0.43, 0.35)
Δ P	0.92	(-0.17, 1.74)
U	0.05	(-0.025, 0.09)

Third state (Volcker)

Responses of Δ M to	Posterior peak estimate	.68 probability interval
Δ Pcom	-0.12	(-0.06, 0.05)
R	0.01	(-0.02, 0.20)
Δ y	0.13	(-0.70, 0.64)
Δ P	0.23	(-0.51, 0.28)
U	0.02	(-0.04, 0.06)

TABLE 12. Annual average output growth rates over 1961-86, actual and counterfactual

Actual	Burns	Greenspan	Volcker
3.7206	4.0560	3.2454	2.8956

TABLE 13. Log marginal data densities for other models

	Excluding money in policy rule		Permanent regime change
	Variances Only	Monetary Policy	Monetary Policy
2 states	13,330.89	13,347.46	13,154.08
3 states	13,432.88	13,419.88	13,414.53
4 states	13,462.40	13,296.58	13,412.85
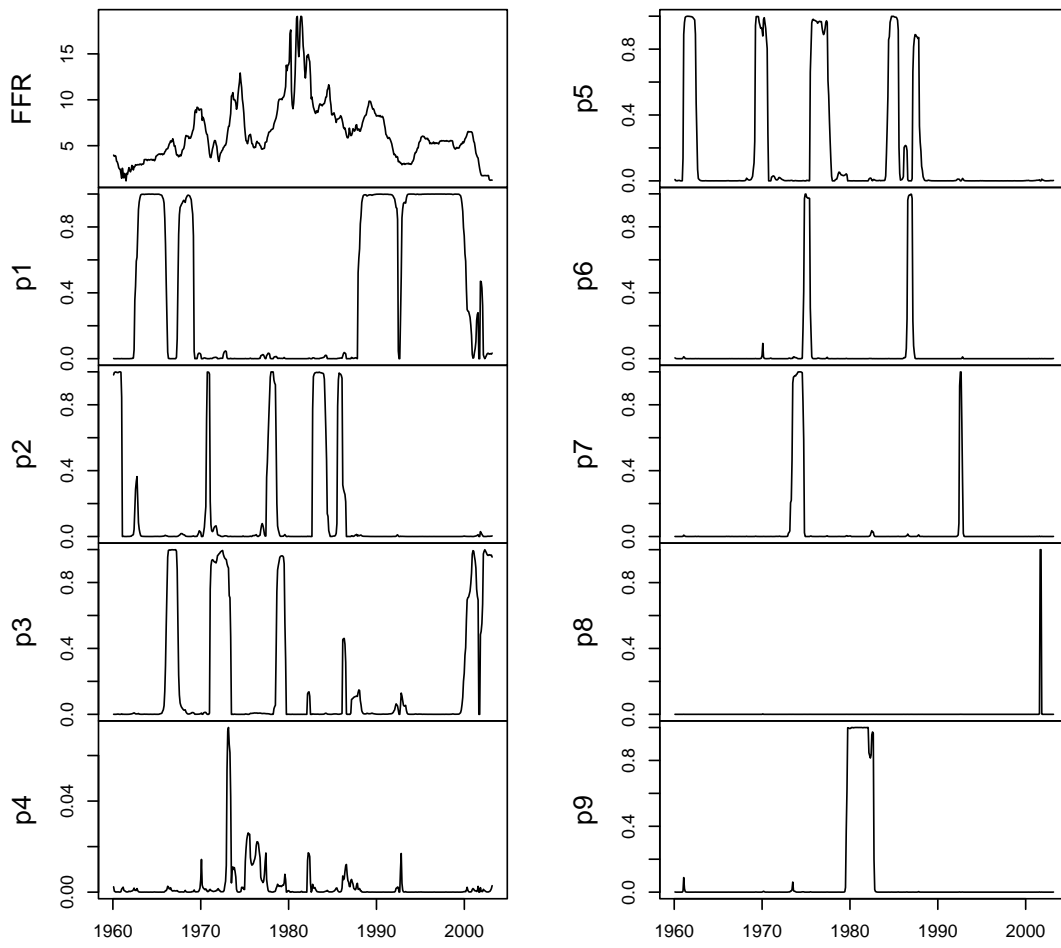
FIGURE 1. 9-state variances-only probabilities; the Fed Funds Rate in upper left.
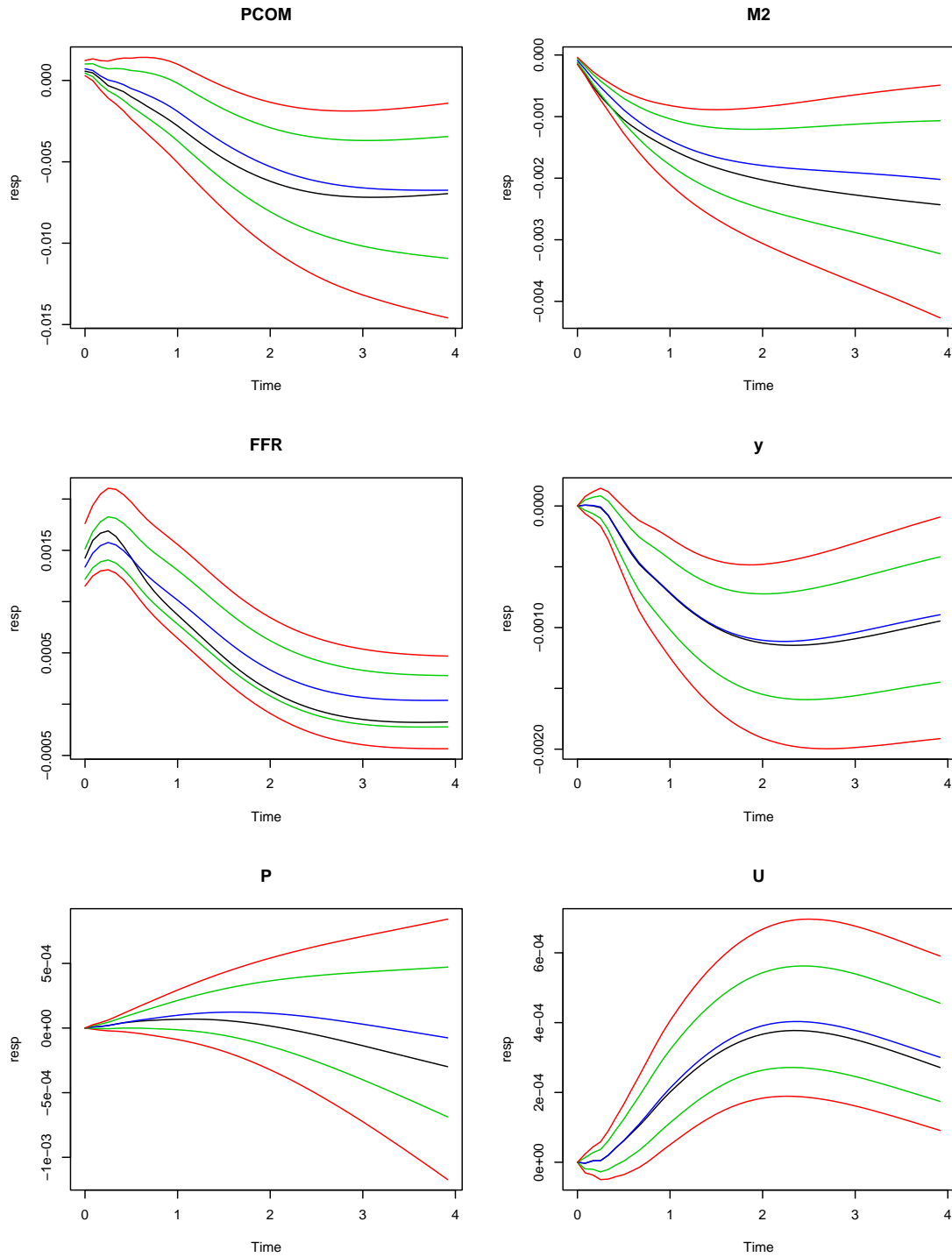
FIGURE 2. Responses to a Monetary Policy Shock, 9-state Variances-Only Model

Note: Each graph shows, over 48 months, the modal's estimated response (blackest), the median response, and 68% and 90% probability bands.
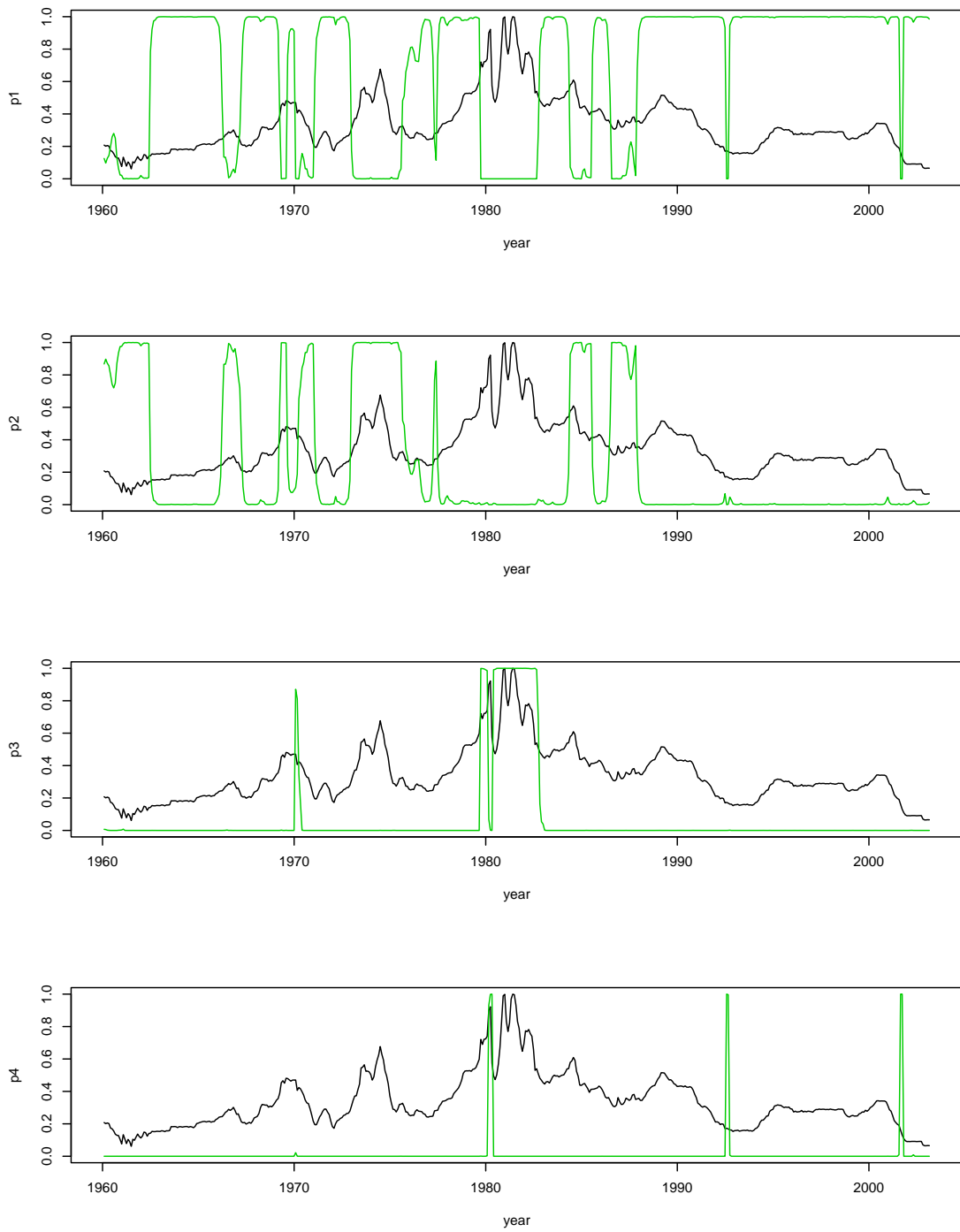
FIGURE 3. State Probabilities, 4-state Monetary Policy Changing

In the background of each figure is the time path of the Fed Funds Rate.
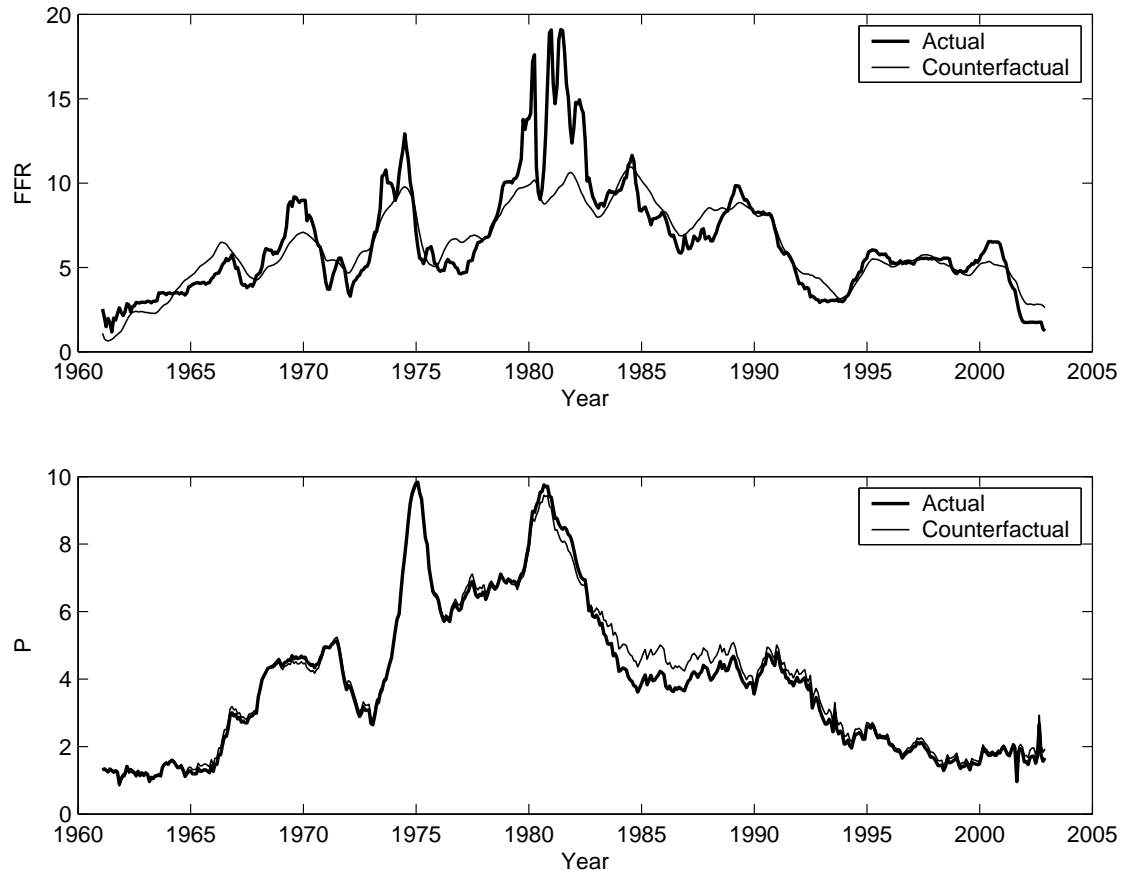
FIGURE 4. Counterfactual paths with no coefficient changes and no policy shocks, 9-state variances-only model
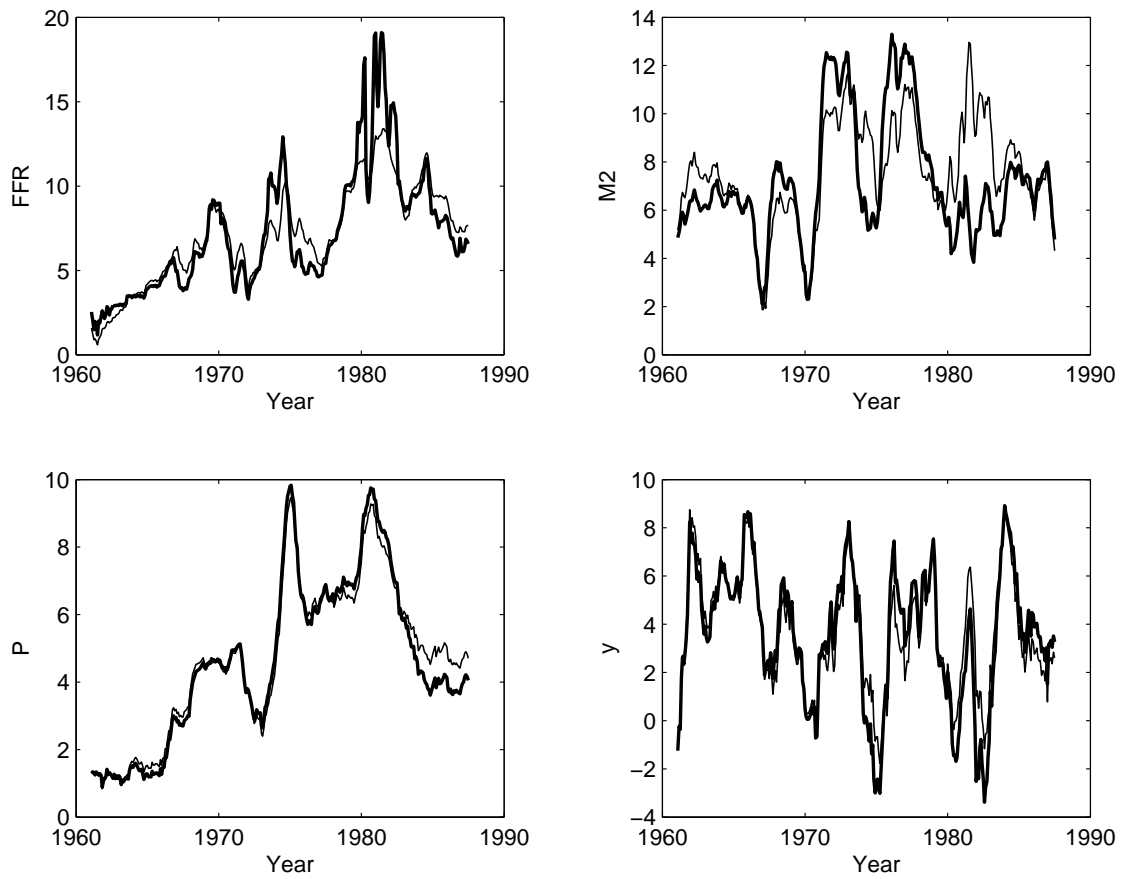
FIGURE 5. Fixed Greenspan policy throughout 1961-87, 4-state monetary policy model. Each graph shows the actual path (thick line) and the counterfactual path (thin line).
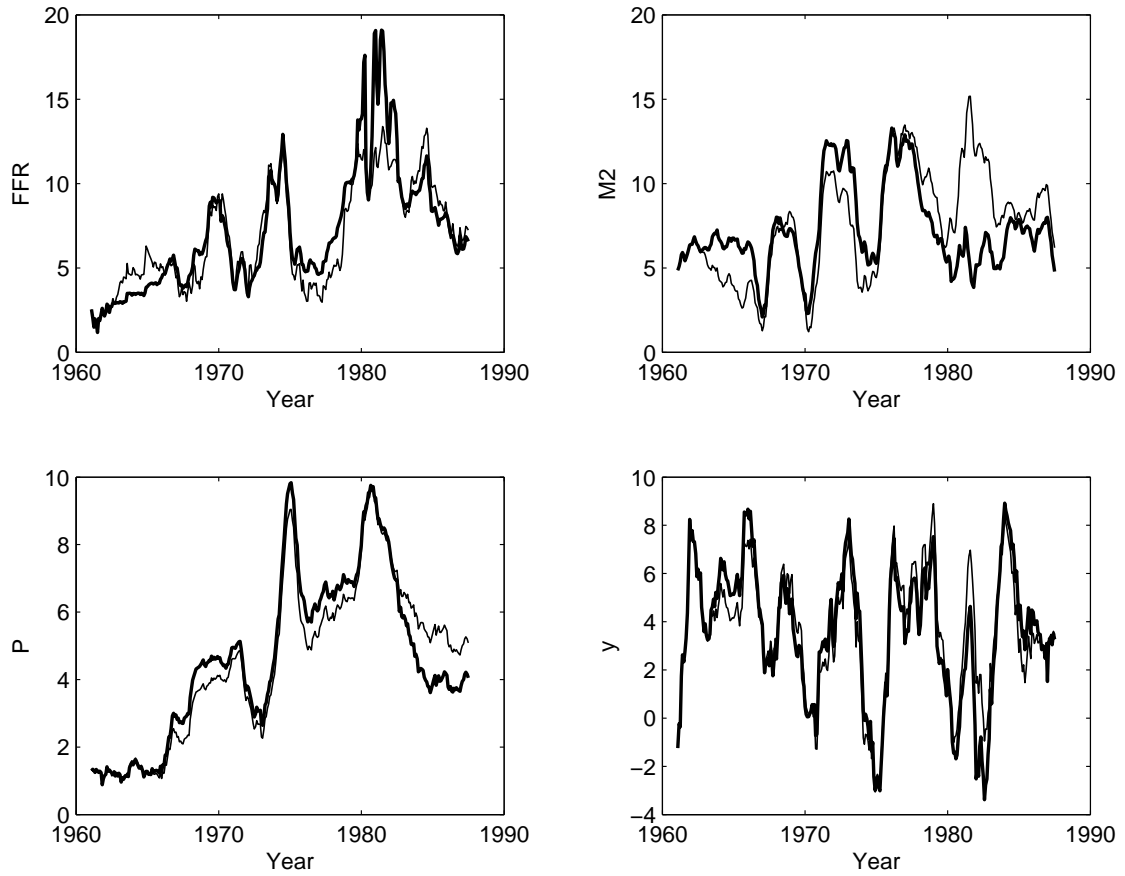
FIGURE 6. Fixed Burns policy throughout 1961-87, 4-state monetary policy model. Each graph shows the actual path (thick line) and the counterfactual path (thin line).
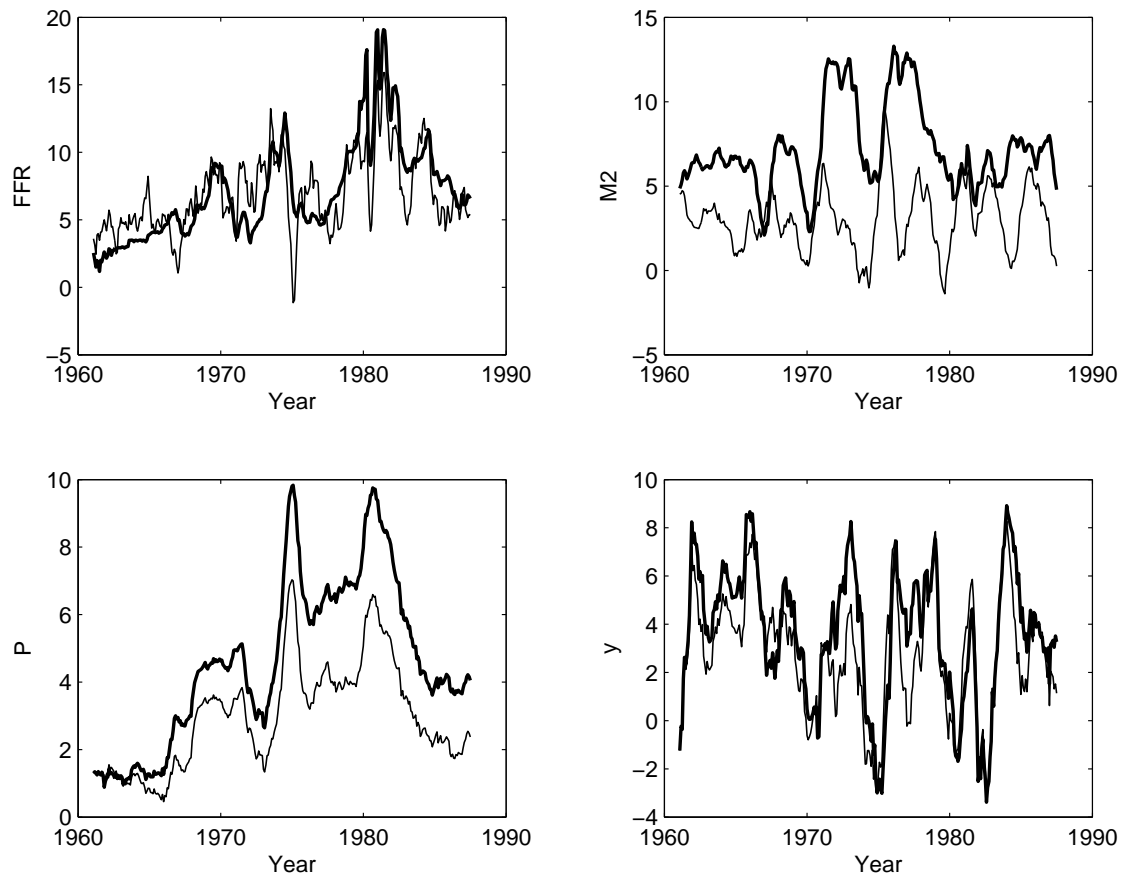
FIGURE 7. Fixed Volcker policy throughout 1961-87, 4-state monetary policy model. Each graph shows the actual path (thick line) and the counterfactual path (thin line).
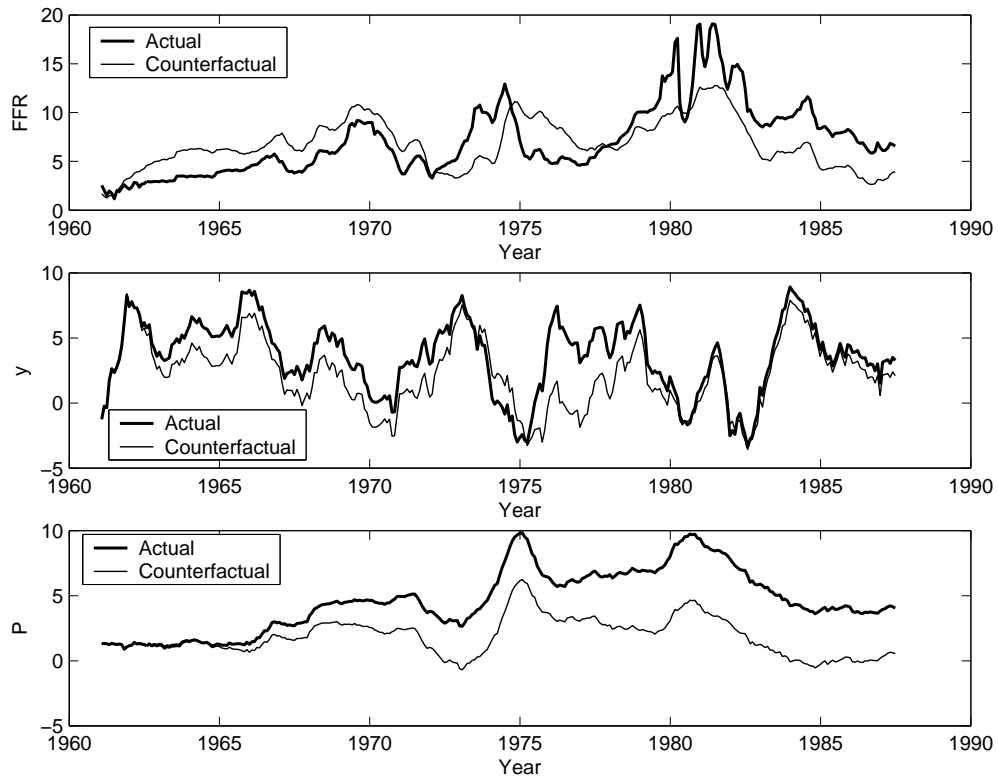
FIGURE 8. Counterfactual paths if policy response to inflation had been twice as strong as in the estimated Greenspan policy throughout 1961-1987, 4-state monetary policy model.
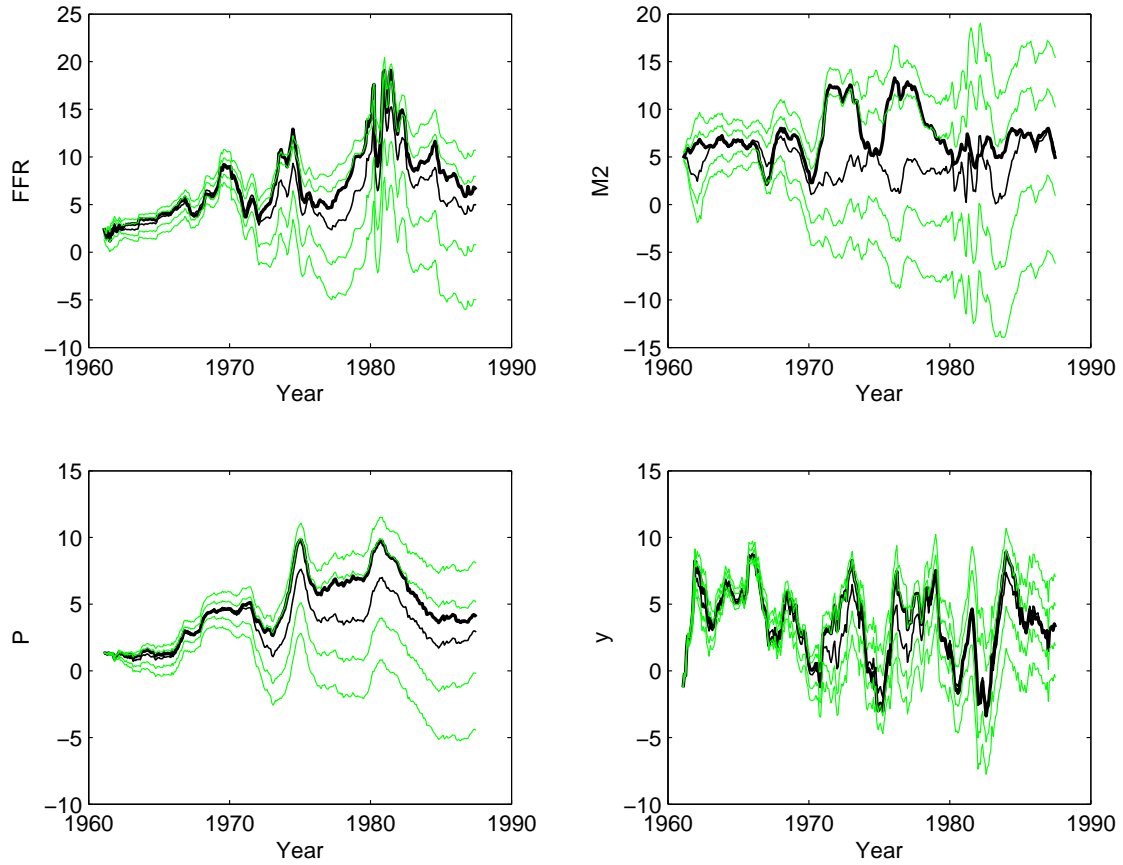
FIGURE 9. The Greenspan policy rule distribution throughout the pre-Greenspan period, 4-state monetary policy model. Each graph shows the actual path (thick line), the median counterfactual path (thin dark line), and the 68% and 90% probability bands (thin light lines).
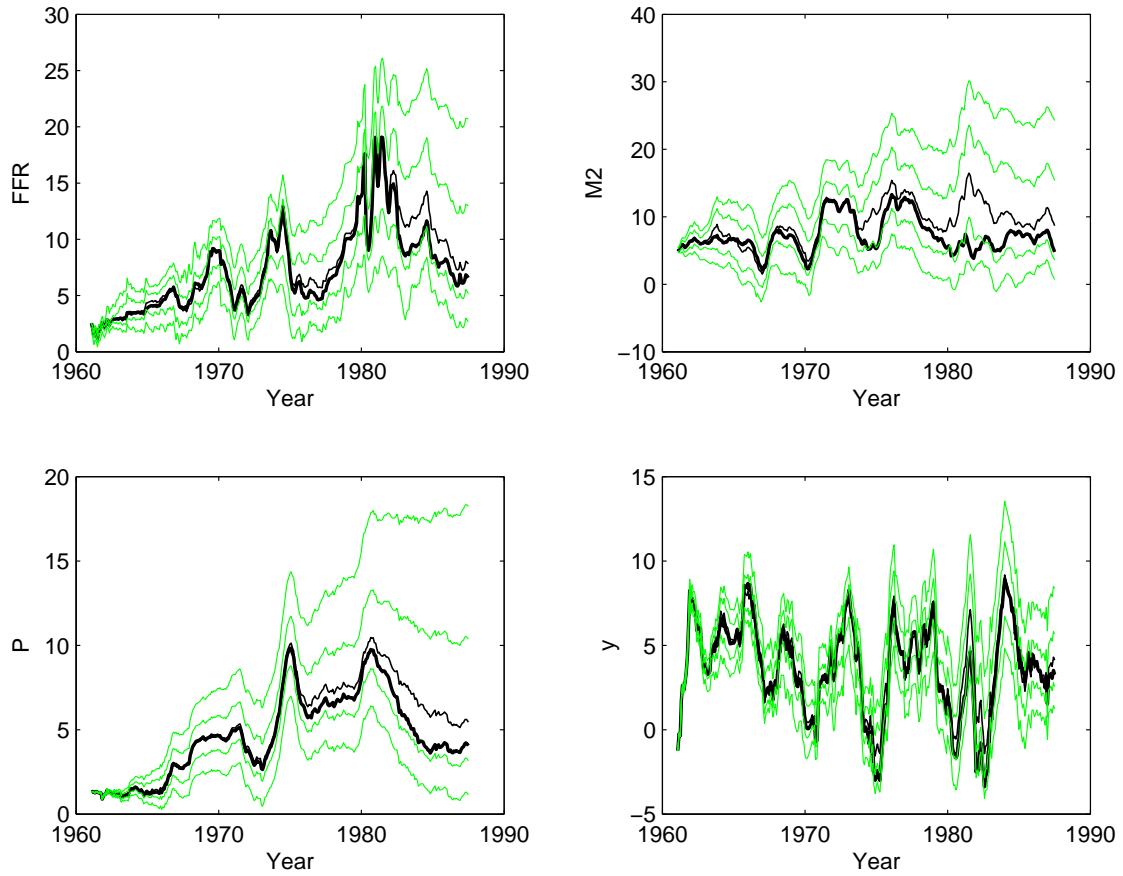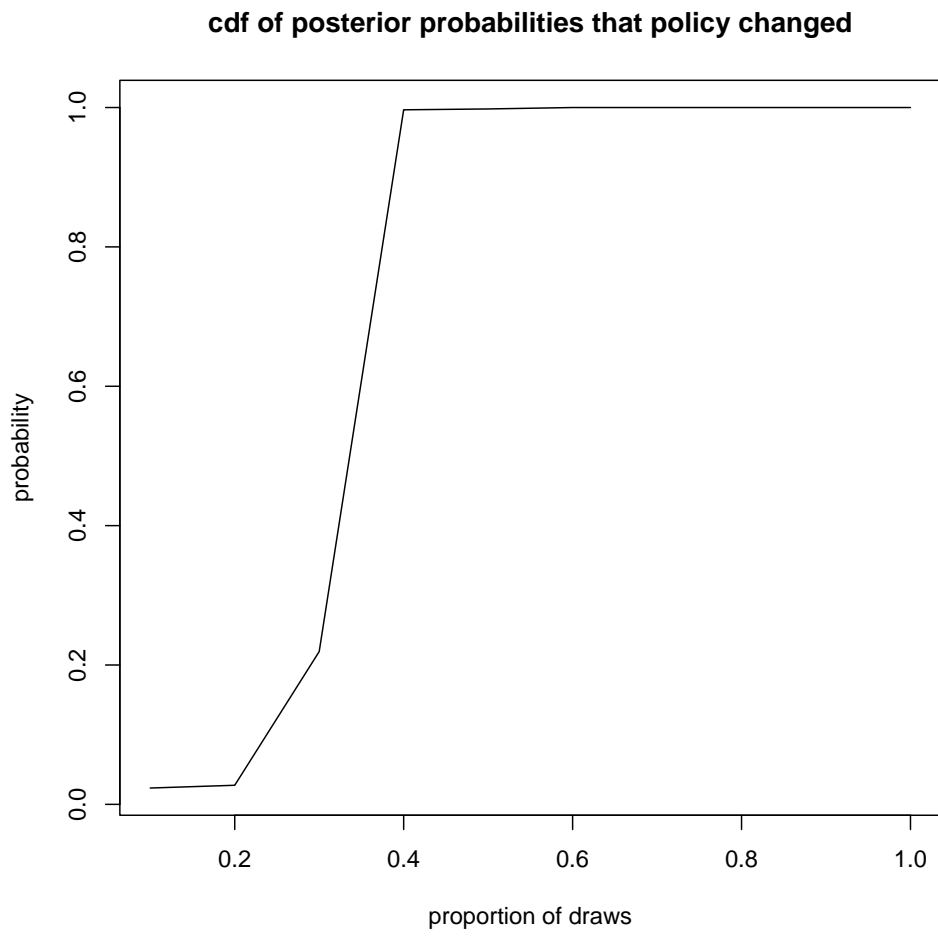
FIGURE 10. The Burns policy rule distribution throughout the pre-Greenspan period, 4-state monetary policy model. Each graph shows the actual path (thick line), the median counterfactual path (thin dark line), and the 68% and 90% probability bands (thin light lines).

cdf of posterior probabilities that policy changed

FIGURE 11. Simulated data: cumulative density function of posterior probabilities that policy changes.

References

Bernanke, Ben and Ilian Mihov, "Measuring Monetary Policy," *Quarterly Journal of Economics*, 1998, *113* (3), 869–902.

Burns, Arthur F., "The Anguish of Central Banking," *Federal Reserve Bulletin*, 1987, *September*, 687–698.

Canova, Fabio and Luca Gambetti, "Structural Changes in the US Economy: Bad Luck or Bad Policy," August 2004. Manuscript, Universitat Pompeu Fabra.

Chappell, Henry W., Rob Roy McGregor, and Todd Vermilyea, *Committee Decisions on Monetary Policy: Evidence From Historical Records of the Federal Open Market Committee*, Cambridge, Massachusetts: The MIT Press, 2005.

Chib, Siddhartha, "Calculating Posterior Distributions and Model Estimates in Markov Mixture Models," *Journal of Econometrics*, 1996, *75*, 79–97.

———— **and Ivan Jeliazkov**, "Marginal Likelihood From the Metropolis-Hastings Output," *Journal of the American Statistical Association*, March 2001, *96* (453), 270–281.

Christiano, Lawrence, Martin Eichenbaum, and Charles Evans, "Nominal Rigidities and the Dynamics Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 2005, *113*, 1–45.

Clarida, Richard, Jordi Gali, and Mark Gertler, "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 2000, *CXV*, 147–180.

Cogley, Timothy and Thomas J. Sargent, "Evolving US Post-Wolrd War II Inflation Dynamics," *NBER Macroeconomics Annual*, 2001, *16*, 331–373.

———— **and** ————, "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.," *Review of Economic Dynamics*, April 2005, *8*, 262–302.

Cooley, Thomas F., Stephen F. LeRoy, and Neil Raymon, "Econometric policy evaluation: Note," *The American Economic Review*, June 1984, *74*, 467–470.

Edge, Rochelle M., "Time to Build, Time to Plan, Habit Persistence, and the Liquidity Effect," International Finance Discussion Paper 673, Federal Reserve Board July 2000.

Gambetti, Luca, Evi Pappa, and Fabio Canova, "The Structural Dynamics of US Output and Inflation: What Explains the Changes?," March 2005. Manuscript, Universitat Pompeu Fabra.

Gelfand, A. E. and D. K. Dey, "Bayesian Model Choice: Asymptotics and Exact Calculations," *Journal of the Royal Statistical Society (Series B)*, 1994, *56*, 501–514.

Geweke, John, "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 1999, *18* (1), 1–73.

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

_____ , **Daniel F. Waggoner, and Tao Zha**, "Normalization in Econometrics," June 2004. Federal Reserve Bank of Atlanta Working Paper 2004-13.

Kim, Chang-Jin and Charles R. Nelson, *State-Space Models with Regime Switching*, London, England and Cambridge, Massachusetts: MIT Press, 1999.

_____ **and** _____ , "Estimation of a Forward-Looking Monetary Policy Rule: A Time-Varying Parameter Model using Ex-Post Data," June 2004. Manuscript, Korea University and University of Washington.

Leeper, Eric M. and Tao Zha, "Modest Policy Interventions," *Journal of Monetary Economics*, 2003, *50* (8), 1673–1700.

_____ , **Christopher A. Sims, and Tao Zha**, "What Does Monetary Policy Do?," *Brookings Papers on Economic Activity*, 1996, *2*, 1 – 78.

Litterman, Robert B., "Forecasting with Bayesian Vector Autoregressions — Five Years of Experience," *Journal of Business and Economic Statistics*, 1986, *4*, 25–38.

Lubik, Thomas A. and Frank Schorfheide, "Testing for Indeterminacy: An Application to U.S. Monetary Policy," *American Economic Review*, 2004, *94* (1), 190–219.

Lucas, Jr. Robert E., "Econometric Testing of the Natural Rate Hypothesis," in Otto Eckstein, ed., *The Econometrics of Price Determination*, Board of Governors of the Federal Reserve and the Social Science Research Council 1972, pp. 50–59.

Primiceri, Giorgio, "Why Inflation Rose and Fell: Policymakers' Beliefs and US Postwar Stabilization Policy," November 2003. Northwestern University.

_____ , "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 2005, *72*, 821–852.

Robertson, John C. and Ellis W. Tallman, "Improving Federal-Funds Rate Forecasts in VAR Models Used for Policy Analysis," *Journal of Business and Economic Statistics*, July 2001, *19* (3), 324–330.

Sargent, Thomas J., Noah Williams, and Tao Zha, "Shocks and Government Beliefs: The Rise and Fall of American Inflation," *The American Economic Review*, in press.

Sims, Christopher A., "A Rational Expectations Framework for Short Run Policy Analysis," in William Barnett and Ken Singleton, eds., *New Approaches to Monetary Economics*, Cambridge University Press 1987, pp. 293–310.

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

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

———, "Implications of Rational Inattention," *Journal of Monetary Economics*, April 2003, *50* (3), 665–690.

——— **and Tao Zha**, "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 1998, *39* (4), 949–968.

——— **and** ———, "Does Monetary Policy Generate Recessions?," *Macroeconomic Dynamics*, in press.

Stock, James H. and Mark W. Watson, "Has the Business Cycles Changed? Evidence and Explanations," 2003. Monetary Policy and Uncertainty: Adapting to a Changing Economy, Federal Reserve Bank of Kansas City Symposium, Jackson Hole, Wyoming, August 28-30.

Waggoner, Daniel F. and Tao Zha, "Likelihood Preserving Normalization in Multiple Equation Models," *Journal of Econometrics*, 2003, *114* (2), 329–347.

FOOTNOTES

[1] In all the models studied here, we incorporate the Robert B. Litterman (1986)'s lag-decay prior that effectively dampens the unreasonable influence of long lags. Thus the over-parameterization problems associated with typical VAR's do not apply here. In addition, the marginal likelihood or the Schwarz criterion used in this paper as a measure of fit, by design, would penalize an excessive number of parameters that over-fit the data.

[2] Of course the expression we give here for the MAR is only valid if the innovations are not stationary infinitely far back into the past, but instead are, e.g., zero before some startup date. Or the expression can be though of as the limit as $\rho \to 1$ of stationary MAR's with coefficients of the form $\left((1 - \rho L)A_0\right)^{-1}$.

[3] Another early paper emphasizing the need for stochastic modeling of policy change is Cooley et al. (1984). More recently Leeper and Zha (2003) have drawn out the implications of this way of thinking for the practice of monetary policy.

[4] As robustness checks, we also used the M2 stock instead of M2 divisia and the CPI (as well as the GDP deflator) instead of the core PCE price index and the paper's main conclusions remained unchanged.

[5] 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 et al., 2005; Edge, 2000; Sims, 2003, 1998). The economic and theoretical justification of the identification presented in Table 1 can also be found in Leeper et al. (1996) and Sims and Zha (in press). This identification has proven to be stable across different sets of variables, different sample periods, and different developed economies.

[6] The reason redundant states create this erratic behavior is that our sampling scheme draws sequences of states as well as parameter vectors. When there is a redundant state, it may be a posteriori likely that the state did not occur at all in the sample. If we have drawn a state sequence in which one state does not occur, at the next Gibbs-sampling step, where coefficients are drawn for each state, the data are exerting no influence at all on the parameters for the non-occurring state; the draw is from the prior for that state's coefficients. Since no draw of the coefficients for that state is likely to improve the likelihood much (the state being redundant), the MCMC chain can remain stuck with no occurrences of the state for a long time. But the same likelihood value can also be achieved with this redundant state occurring, but having coefficients

matching those of some other state. Once such a draw occurs, the transition probabilities between the two nearly identical states become ill determined by the likelihood. This leads to the MCMC chain being stuck for a long time in a mode with all states occurring, but one or more states nearly identical. All of these draws are likely to imply low posterior density values, but they can vary quite a bit, depending on the shape of the prior. Thus we obtain MCMC chains with very strong serial correlation, making estimates of sample means unreliable, but still allowing a reliable conclusion that posterior weight on the model is low.

[7]Note, though, that the "private sector" and "all change" models may be doing less well because of parameter count. It could be that more tightly parameterized models of coefficient change in the private sector would look better in a table like this.

[8]The shape of the impulse responses as seen on scaled plots is the same across states.

[9]Note that if we calculated long run responses of the interest rate for this regime, variable by variable, we would get very large, opposite-signed numbers that would have high uncertainty and be difficult to interpret.

[10]Computations for these simulated data are quite intensive. For each model, it takes about a week on a single processor computer to get the marginal data density. There are a total of 40 models (which would be a 10-month computation). We acknowledge the technical support from the College of Computer Science at the Georgia Institute of Technology, which designed a Linux-based program called "STAMPEDE" specifically for this project. This program allows us to run our jobs efficiently on a cluster of computers simultaneously.

[11]To obtain an absorbing state, our original prior on the transition matrix is modified so that the Dirichlet weight α_{kk} on the diagonal element of the transition matrix is 1.0. The original prior gives the weight value of 5.0, which effectively puts an upper bound on the estimate of p_{kk} away from 1.0. In this case, we obtain the posterior probabilities of this state being near one for almost the entire period for which the state actually prevails.

[12]The Schwarz, or Bayesian Information, Criterion, is usually described as log likelihood minus number of parameters times log of sample size divided by 2. Under standard regularity conditions it is guaranteed to be maximal at the model with highest posterior odds, if the sample is large enough. Though here we use LPD in place of LLH, the same asymptotic reasoning that justifies the criterion based on likelihood applies here.

[13]Indeed, a tighter prior on $\lambda_{ij}(k)$ tends to lower the marginal likelihood for the same model.

[14]We are still improving our algorithm. Once it is finished, it is possible that the computing time could be considerably reduced.

[15]Note that the marginal data density is invariant to the way parameters are normalized, as long as the Jacobian transformations of the parameters are taken into account explicitly.

[16]This label normalization is a computationally efficient way to approximate Wald normalization discussed by Hamilton et al. (2004).