WERE THERE REGIME SWITCHES IN US MONETARY POLICY?

CHRISTOPHER A. SIMS AND TAO ZHA

ABSTRACT. A multivariate model, identifying monetary policy and allowing for simultaneity and regime switching in coefficients and variances, is confronted with US data since 1959. The best fit is with a model that allows time variation in structural disturbance variances only. Among models that allow for changes in equation coefficients also, the best fit is for a model that allows coefficients to change only in the monetary policy rule. That model allows switching among three main regimes and one rarely and briefly occurring regime. The three main regimes correspond roughly to periods when most observers believe that monetary policy actually differed, and the differences in policy behavior are substantively interesting, though statistically ill-determined. The estimates imply monetary targeting was central in the early 80's, but also important sporadically in the 70's. The changes in regime were essential neither to the rise in inflation in the 70's nor to its decline in the 80's.

I. THE DEBATE OVER MONETARY POLICY CHANGE

In an influential paper, Clarida, Galí and Gertler 2000 (CGG) presented evidence that US monetary policy changed between the 1970's and the 1980's, indeed that in the 70's it was drastically worse. They found that the policy rule apparently followed in the 70's was one that, when embedded in most 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

Date: April 21, 2004.

 We are indebted to Dan Waggoner for many helpful discussions as well as his invaluable help on C programming. Eric Wang provided excellent research assistance in computation on the Linux operating system.

other hand, eliminated this indeterminacy. These results are a possible explanation of the volatile and rising inflation of the 70's and of its subsequent decline.

The CGG analysis has two important weaknesses. One is that it fails to account for stochastic volatility. US macroeconomic variables, and particularly the federal funds rate, have gone through periods of tranquility and of agitation, with forecast error variances varying greatly from period to period. Ignoring such variation does not lead to inconsistent estimates of model parameters when the forecasting equations themselves are constant, but it strongly biases — toward a finding of changed parameters — tests of the stability of the forecasting equations.

The other weakness is that the CGG analysis rests on powerful and implausible identifying assumptions. They require that we accept that the response of the monetary authority to expected future inflation and output does not depend on the recent history of inflation, money growth, or output. It is hard to understand why this should be so, especially in the 70's, when monetarism was a prominent theme in policy debates, Congress was requiring reports from the Fed of projected time paths of monetary aggregates, and financial markets were reacting sensitively to weekly money supply numbers. The requirement for existence and uniqueness of equilibrium in dynamic models is that the monetary policy rule show a more than unit response of interest rates to the sum of the logs of all nominal variables that appear on the right-hand side of the reaction function. If we force a particular measure of expected future inflation to proxy for all the nominal variables that actually appear independently in the reaction function, we are bound to get distorted conclusions. On the one hand, because expected future inflation will be a "noisy" measure of the full set of nominal influences on policy, we might get downward bias in our estimates from the usual errors-in-variables effect. On the other hand, to the extent that expected future inflation (like most expected future values) shows less variation than current nominal variables, we could find a mistaken scaling up of coefficients.

It should not be surprising that the CGG approach is fragile. For one thing, most equilibrium models contain something like the Fisher equation, relating the nominal rate to expected inflation and the real rate. The usual sort of discussion of whether instrumental variables are uncorrelated with disturbances does not confront the question of what creates a clear distinction between, on the one hand, the "forward-looking Taylor rule" equation relating expected future inflation and a real variable to the current nominal rate and, on the other hand, the "Fisher equation" that also relates expected future inflation and a real variable to the current nominal rate. Indeed in some simple models the real rate is determined by expected future output growth, which might be a candidate as the expected real variable on the right-hand side variable of a forward-looking Taylor rule. If the distinction between these equations is statistically weak, no amount of testing of overidentifying restrictions will detect the problem — both equations will satisfy all the exclusion restrictions.

If the actual policy rule is at least partly backward-looking, the CGG identifying restrictions can produce very misleading results. For example, it is plausible that when monetary policy raises the interest rate, it tends to bring down inflation, with a delay. In other words, we can easily imagine an economy in which, when we observe high interest rates, we conclude that monetary policy is tight and that therefore inflation will soon be low. This would certainly not imply that monetary policy is allowing non-uniqueness of equilibrium, but it would imply that the partial correlation of current interest rates with future inflation is negative. And of course a small component of this effect, combined with the Fisher relation, could easily produce a positive partial correlation but a CGG coefficient on expected inflation less than one.

A simple example of a New Keynesian model with a unique equilibrium, but in which the CGG methods would produce misleading results, appears in Appendix A.

Bernanke and Mihov (1998) made their identifying assumptions explicit in a multivariate model, and they concluded that there was little evidence of major shifts in monetary policy. Like the rest of the structural VAR literature, (and unlike users of the single-equation CGG setup) they validate their identifying assumptions by displaying impulse response functions that let us assess whether their estimates imply plausible policy effects in a full dynamic system. They did not model time-varying volatility, however.

Cogley and Sargent (2002) apply the CGG identifying assumptions in the context of a model with stochastically drifting parameters and stochastic volatility. Though they emphasize that their results are consistent with their own earlier work that ignored stochastic volatility and found important changes in policy parameters, after accounting for stochastic volatility their results are also largely consistent with the hypothesis that there has been no drift at all in the parameters of the policy rule. This latter result is in sharp contrast with their earlier work that ignored stochastic volatility. It is disappointing that, despite having an estimated multivariate model available, they did not check whether their identifying assumptions are consistent with plausible impulse response to monetary policy disturbances.

Boivin and Giannoni (2003) test for structural change in a VAR model and find strong evidence of a change using an asymptotically justified hypothesis test. They do not include a monetary aggregate in their system. They describe the test they use as "heteroskedasticity-consistent", but it allows for changing variances in only a limited sense. It accounts for fluctuating disturbance variance that can be explained by right-hand-side variables (through use of heteroskedasticity-consistent covariance matrix estimators) and it also allows for a single change in residual variances somewhere in the sample (not necessarily synchronized with the break in coefficient values that is being tested for). Eyeball inspection of forecast errors for short interest rates before, during, and after the 1979-82 period makes it evident that a single shift in variance will not capture the actual historical experience, and it is hard to argue that the period of increased variance in interest rate disturbances was predictable from right-hand-side variables in a VAR. The conclusion in this Boivin-Giannoni paper that there is coefficient change therefore reflects the same sort of bias as do other papers that do not make realistic allowance for changing disturbance variances.

Boivin (2004) is a single-equation study using a version of the CGG identification assumptions. It finds evidence of coefficient change, using a test statistic with the same weaknesses as that in Boivin and Giannoni (2003). Its estimated time patterns of changes in the policy rule do agree qualitatively with the findings of this paper, though, in that it finds large changes that are later reversed, rather than a monotonic evolution.

Primiceri (2003b,a) studies changes in monetary policy in two papers. In one he provides a tightly parameterized model that he uses to interpret the historical record as reflecting learning by monetary policy makers about the structure of the economy. In the other he models time variation along lines similar to Cogley and Sargent, but with more complete and explicit treatment of identification. In the latter paper he concludes that there has been time variation in US monetary policy, but that it has not been of great quantitative importance. The paper on learning appears in conflict with the other paper, since it argues that the rise and fall in US inflation can be explained by policy-makers' learning. The conflict is not necessarily strong, however. The work of Cogley and Sargent and Primiceri all fits with the notion that the data do not deliver clear evidence of parameter change unless one imposes strong, and potentially controversial, overidentifying assumptions — which is exactly what Primiceri does in his learning paper.

This paper follows Primiceri in using a multivariate model with stochastic volatility and explicit identifying assumptions that allow us to consider monetary policy change. Unlike Primiceri, we model parameter change as discrete, discontinuous, stochastically timed, changes in parameters and variances. The type of model we use is known as a "Markov switching" or "hidden Markov chain" model. This approach seems well suited to the period we study, because the "Volcker reserves targeting" period, October 1979 through 1982, is widely recognized to have constituted a sudden shift to a new pattern of policy behavior. The models used by Primiceri and Cogley and Sargent imply a more nearly continuous time path for parameters and therefore do not track well around the October 1979 date. Also, the Markov switching framework includes as a limiting case a model of fat-tailed distributions for disturbances, treated as mixtures of normal random variables with different variances. Finally, the theoretical rational expectations monetary policy literature has often argued that the only important policy changes are "regime shifts", which are modeled as once-and-for-all discontinuous changes. It is therefore of some interest to use a modeling framework that could detect such regime shifts if they did occur historically in US data.

Our conclusions have two levels. The most important result is simple: 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. The Volcker reserves targeting period then emerges simply as a period of high variance in disturbances of the policy rule. However, like Cogley and Sargent and Primiceri, we find that if we do allow coefficients to change, the point estimates of the changes are not substantively trivial, even though the data leave their magnitudes uncertain.

So a second level of our analysis explores the best-fitting model we have found that does allow change in parameters other than structural equation variances. That model is one that allows the strength of monetary policy responses to vary with the regime, but with other parameters remaining fixed, except for equation variances. 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 third 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 third 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.) For all of the three regimes our estimates imply that with high probability monetary policy responses to inflation were strong enough to guarantee a determinate equilibrium price level.

We display counterfactual simulations of history with alternate monetary policy regimes. Any one of the three main regimes could have been held in place, we conclude, and the pattern of rising inflation in the 70's, followed by decline in the 80's, would with high probability have been maintained. The steepness of the rise and of the fall in inflation would have been different under different regimes, as would the depth of recessions that occurred along the way.

The model implies that Greenspan's monetary policy, had it been in place through the 70's, would have resulted in a less steep rise in inflation than actually occurred — but that it would have done so without (with high probability) at any point pushing interest rates higher. Apparently the model's dynamics imply that the harsher stop-go pattern of actual monetary policy in the 70's, with deeper recessions and very rapid expansions on emergence from recession, was more responsible for the rise in inflation than was any general tendency to keep interest rates low.

We think these empirical results have important 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.
- We should look for structural modeling ideas that might match the observation that stop-go policies can generate rising inflation even with high average interest rates.
- 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.

II. Class of Models

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

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

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

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.

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:

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

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 Hamilton 1989 and Chib 1996, we impose no restrictions on the transition matrix $P = [p_{ik}]$. [1]

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

$$y_t' = x_t' B(s_t) + u'(s_t), \quad t = 1, \ldots, T; \tag{3}$$

where

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

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

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

---

[1] As shown in Sims and Zha 2004, the class of multiple equation models considered in this paper is complex and pushes the limits of what our computers and analytical capacity can handle. One could in principle (and we intend to do so in future work) give special structure to P to investigate a variety of models of parameter change. For example, one could create two classes of state variable, one indexing variances and one indexing equation coefficients. These could be allowed to evolve independently. If the transition matrices for the two types of state are Q_1 and Q_2, we get the desired independent evolution by setting $P = Q_1 \otimes Q_2$. We could also postulate an s that takes on many values, but with the values interpreted as subsets of the plane over which the joint distribution of x_t and x_{t-1} from an autoregressive model are spread. P's entries are then filled in as the conditional probabilities of these subsets of the plane. This would allow us to have a large P whose entries are functions of a small number of parameters, and to approximate arbitrarily well the kind of smooth drift in parameters assumed by Cogley and Sargent and by Primiceri. But in this framework we could easily allow for occasional discontinuous jumps as well as smoother drift. Moreover, a restriction that parameter change is monotonic, with no state recurring after it has been exited, can be implemented by requiring that P be upper triangular.

In the reduced form (4)-(6), $B(s_t)$ and $u(s_t)$ 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 1998a 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.

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

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

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 1998a). 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.

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

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

where $\xi_j(s_t)$ is a scale factor for the j^{th} structural equation, $a_{0,j}(s_t)$ is the j^{th} column of $A_0(s_t)$, $d_j(s_t)$ is the j^{th} column of $D(s_t)$, $d_{ij,\ell}(s_t)$ is the element of $d_j(s_t)$ for the i^{th} variable at the ℓ^{th} 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 coeffcents 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 Lucas (1972).

However, as one of us has explained at more length elsewhere (Sims, 1987), 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.[2]

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.[3]

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

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

[3]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.

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

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.[4]

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 B). 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 coefficienta and impulse responses more stable across different sample periods.

We compare five types of models

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

[4]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 (e.g., Christiano, Eichenbaum, and Evans 2001, Edge 2000, Sims 2003; 1998). The economic and theoretical justification of the identification presented in Table 1 can also be found in Leeper, Sims, and Zha 1996 and Sims and Zha 1998b. This identification has proven to be stable across different sets of variables, different sample periods, and different developed economies.

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	$13,466.86$	$13,480.18$	DEG	DEG
5 states	$13,455.26$	$13,400.10$	DEG	DEG
6 states	$13,510.31$	DEG	DEG	DEG
7 states	$13,530.71$	DEG	DEG	DEG
8 states	$13,540.32$	DEG	DEG	DEG
9 states	$13,544.07$	DEG	DEG	DEG
10 states	$13,538.03$	DEG	DEG	DEG

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 B 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 we were unable to obtain convergence when the number of states became too large and label these situations by "DEG."[5] 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 error in these numbers (based on two million MCMC draws) is from ± 2 to ± 4. For the lower boundary in each column the error is larger. These estimates of MCMC 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.[6]

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

[5] DEG stands for "degenerate." Models with large number of parameters overfit to the data and consequently some states become redundant. These states are not drawn at all in our Markov chain Monte Carlo (MCMC) simulations. In some cases, such degenerate draws led to much lower values of log likelihood but these values can fluctuate wildly from one sequence of MCMC draws to another.

[6] Note, though, that because models with too many parameters are clearly paying a penalty here, it may be 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.

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

0.9643	0.0063	0.0117	0.0064	0.0108
0.0030	0.9394	0.0047	0.0070	0.0210
0.0104	0.0159	0.9455	0.0064	0.0046
0.0026	0.0043	0.0042	0.9476	0.0040
0.0058	0.0155	0.0044	0.0068	0.9425
0.0027	0.0056	0.0058	0.0064	0.0051
0.0052	0.0042	0.0081	0.0068	0.0040
0.0033	0.0041	0.0069	0.0062	0.0038
0.0026	0.0046	0.0087	0.0065	0.0042

0.0057	0.0107	0.0095	0.0049
0.0062	0.0061	0.0069	0.0112
0.0063	0.0064	0.0096	0.0057
0.0058	0.0056	0.0062	0.0051
0.0185	0.0058	0.0064	0.0057
0.9406	0.0120	0.0062	0.0050
0.0057	0.9423	0.0062	0.0053
0.0056	0.0054	0.9429	0.0049
0.0056	0.0056	0.0062	0.9522

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 1st 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

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

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 to the interest rate than the money stock in response to the economic development.

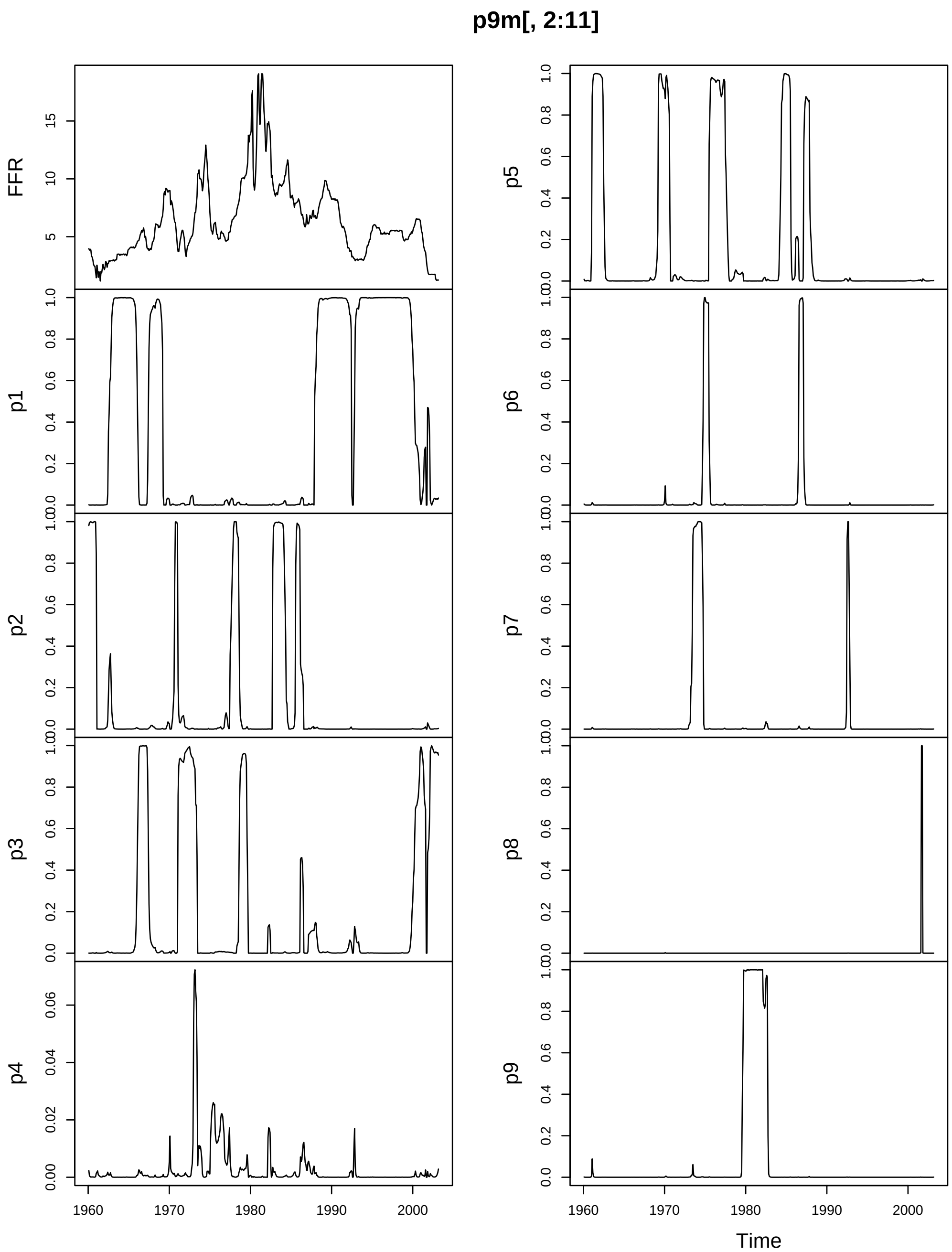

FIGURE 1. 9-state variances-only probabilities; the Fed Funds Rate in upper left.

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.[7] 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 serveral 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 ℓ^{th} lag of the "right-hand-side" variable and δ_ℓ is the coefficient on the ℓ^{th} 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 $\Delta \mathsf{y}$ and $\Delta \mathsf{P}$ 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 $\Delta \mathsf{PCom}$, $\Delta \mathsf{M}$, and $\Delta \mathsf{P}$) 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

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

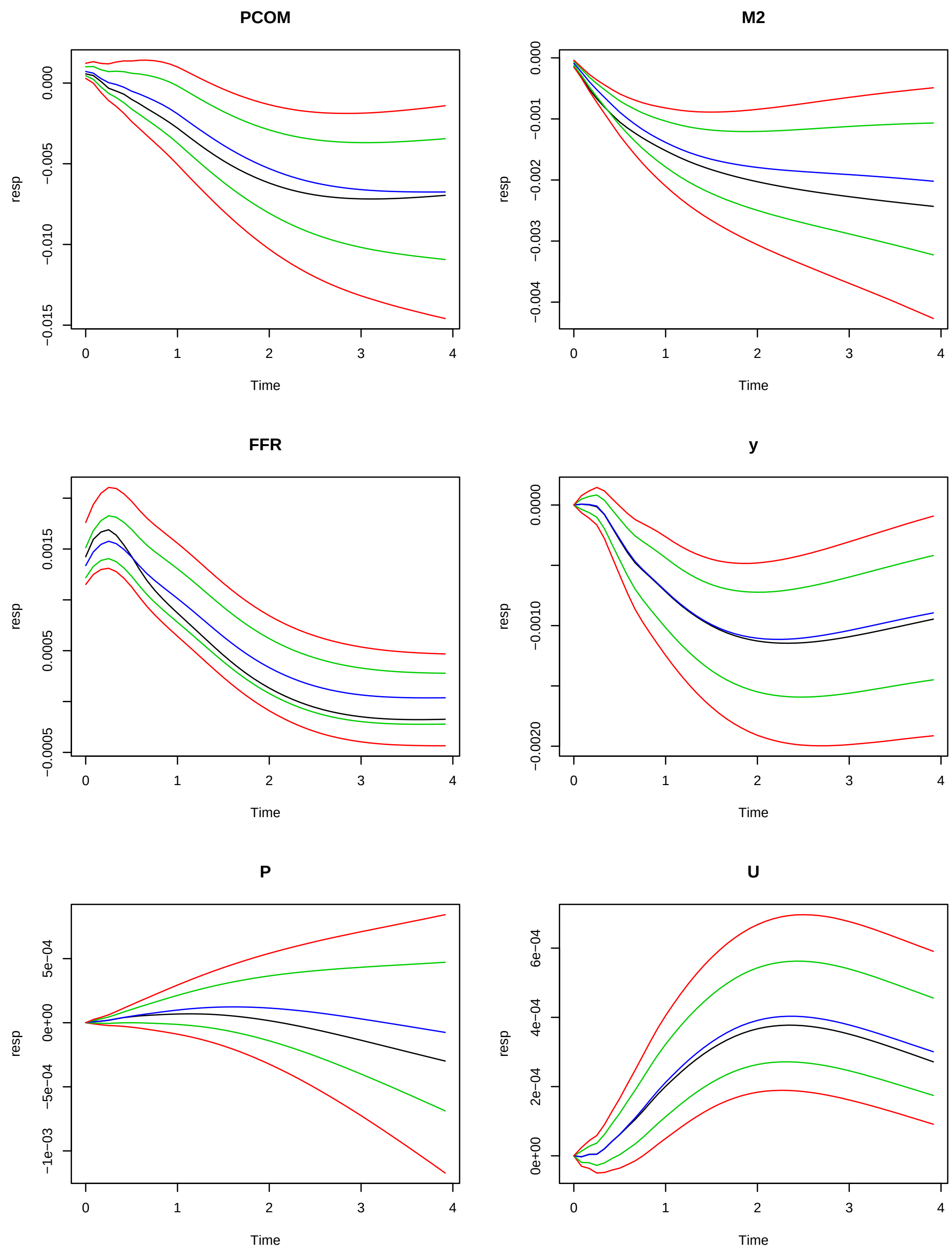

FIGURE 2. Responses to a Monetary Policy Shock, 9-state Variances-Only Model

Note: Each graph shows, over 48 months, the modal's estimated response (blackest), the median response, and 68% and 90% probability bands.

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)

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, occuring most frequently from the early 60's through the early 70's (the first oil shock period),

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

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 a 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 and Chappell and McGregor 2000) and that Greenspan made the interest rate the explicit policy instrument.

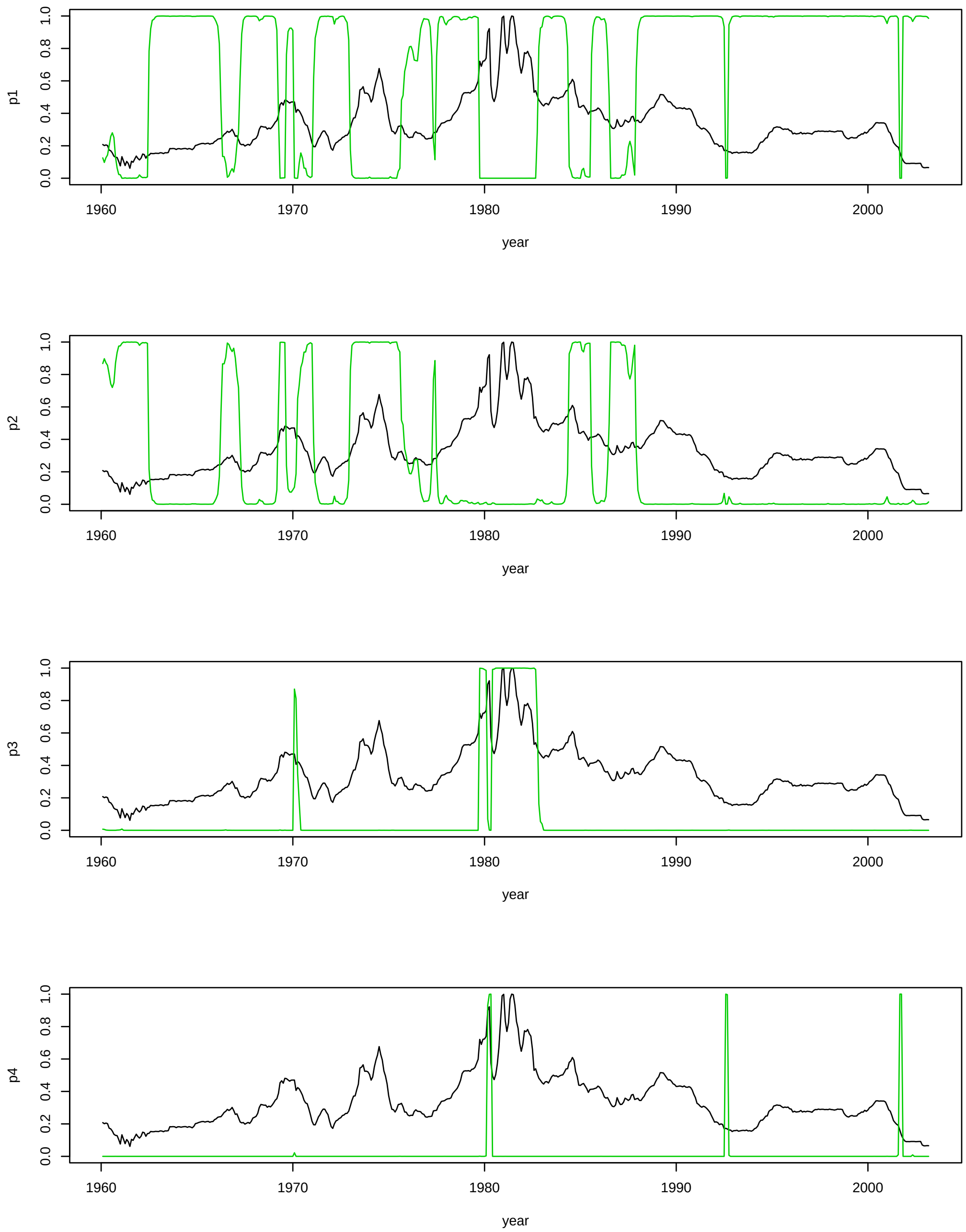

FIGURE 3. State Probabilities, 4-state Monetary Policy Changing

In the background of each figure is the time path of the Fed Funds Rate.

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

The long run policy responses to macro variables show the 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

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

makes money unresponsive to all variables (measured by both point estimates and error bands). The Burns regime shows a disturbingly high responsiveness 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.[8] 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

[8]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.

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)

is that the estimated policy changes do make a noticeable difference, but not a drastic difference. In the following we display three examples that seem most relevant to the debate on the effects of monetary policy changes.

VI.1. **Replacing Volcker regime by Burns regime.** Our estimated Burns regime, which corresponds to only some years of Burns chairmanship when money growth was probably playing some role in policy making, shows a fairly high responsiveness of money growth to inflation in the estimated policy reaction function. It is therefore interesting to see what would have happened to the economy had this regime prevailed in the early 80's. To conduct this exercise, we hold the Burns regime in place through the entire period of Volcker chairmanship. Our simulations account for uncertainty in the coefficients of the estimated Burns regime and uncertainty about the actual historical structural shocks. They keep the unscaled shocks at their historical values (subject to the uncertainty), but scale them to match the Burns regime variances. This makes sense for monetary policy shocks. If one believed, as some policy makers do and the Lucas critique suggests, that private sector shock variances responded to changes in the monetary policy rule, then our use of rescaled private sector shocks would be appropriate. [9]

We see from Figure 4 that as many would have expected, the Burns regime would not have pushed interest rates so high, would have dampened the recession of the early 80's — possibly even to the point where it would not have registered as an official recession — and would have run a substantial risk of letting inflation remain at 8% in the late 80's.[10] The risk of higher inflation, however, is ill-determined, given the wide error bands. Moreover, the most likely path with the Burns regime, as often looked at in the literature, is remarkably close to the historical path. The median outcome is estimated to be about one percentage point higher than historical inflation at the end of the 80's, while having kept unemployment one to two percentage points lower during a long stretch of the early 80's. It is not so obvious that this is terrible monetary policy. The reason the most likely counterfactual path of inflation shows a steady decline from about 9% to 5% is partly due to the result that the response of the interest rate to all nominal variables under the Burns regime is well above

[9]The results we display here are not heavily dependent on shifts in private sector shock variances across regimes. If we replace the Volcker policy rule by the Burns reaction function but keep private shock variances unchanged for this period, the estimated path and error bands are quite similar to Figure 4. One may think that error bands should be narrower if only the policy rule is replaced, but statistically the error bands may be narrower or wider, depending on how the replaced policy reaction function interact with the coefficients and historical shocks in the rest of the system and how informative our normalization on labeling states is.

[10]These simulated results, as well as those in next two sections, are generated with at least 1 million MC draws.

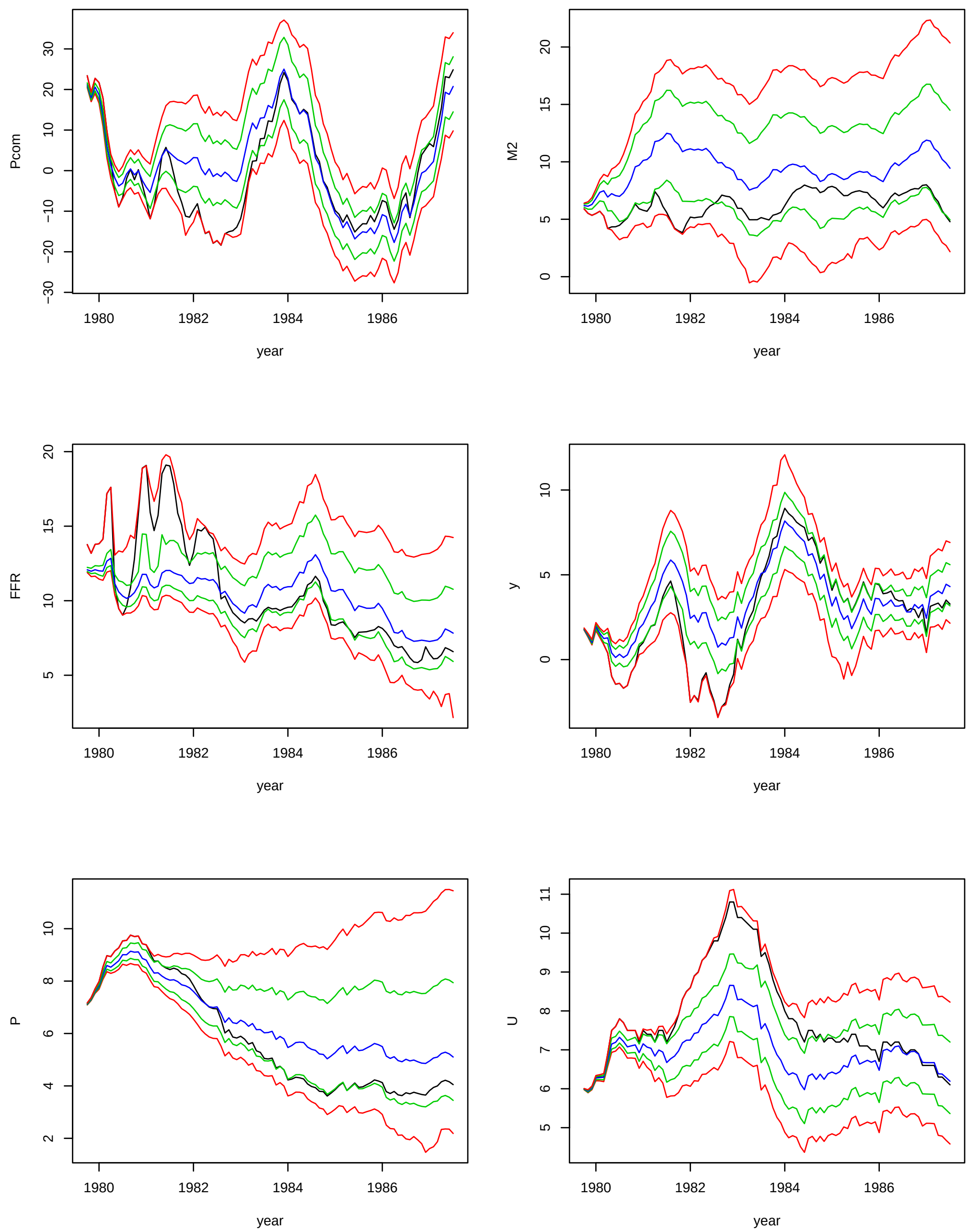

FIGURE 4. Counterfactuals: Burns regime through the 80's

Each graph shows the historical path (blackest), the median counterfactual path, and 68% and 90% probability bands.

1.0 with about 84% probability. Money plays an indispensable role so that any structural model should include the money stock to study the policy rule for this historical period.

VI.2. **Replacing Volcker regime by Greenspan regime.** Towards the end of the 70's, inflation began to gain momentum to rise again. Burns himself recognized the public and political pressure to do something drastic that differed from the previous policy.[11] He speculated as to whether central bankers, "having by now become accustomed to gradualism, would be willing to risk the painful economic adjustments that I fear are ultimately unavoidable" (Burns 1987), but he himself was by then apparently willing to risk it.

But was the drastic change to the reserve-targeting policy adopted by Volcker from 1979 to 1982 in fact necessary to bring down the rapidly rising inflation? After all, Burns's speech, after he left office, suggests that he himself would have made sharply restrictive moves had he stayed in office, and it is his behavior that the model allocates to the Burns regime. One way to answer this question is to rerun economic history, replacing the 1979-1982 Volcker regime with the Greenspan regime, which used an interest rate instrument and smoothed interest rates, thereby being perhaps more "gradualist" than Volcker during 1979-82. We hold the Greenspan regime in place through the entire period of Volcker chairmanship until 1987:7. Would this change in policy have greatly affected outcomes? Would inflation have taken much longer to end?

The simulated results are reported in Figure 5. The counterfactual funds rate path is much smoother than both actual data and the counterfactual path with the Burns regime in Figure 4: the funds rate would have been much lower from late 1979 to the end of 1982 (measured by both the median path and error bands) and the most likely path would have remained lower after 1982. The wide error band for the counterfactual path of money growth is consistent with the less attention paid to money in the estimated policy rule under the Greenspan regime. The counterfactual inflation path would have come down as steadily

[11] In his 30 September 1979 speech (Burns 1987), Burns admitted: "In the United States a great majority of the public now regard inflation as the Number One problem facing the country, and this judgment is accepted by both the Congress and the Executive Branch. ... In view of the strong and widespread expectations of inflation that prevail at present, I have therefore reluctantly come to believe that fairly drastic therapy will be needed to turn inflationary psychology around."

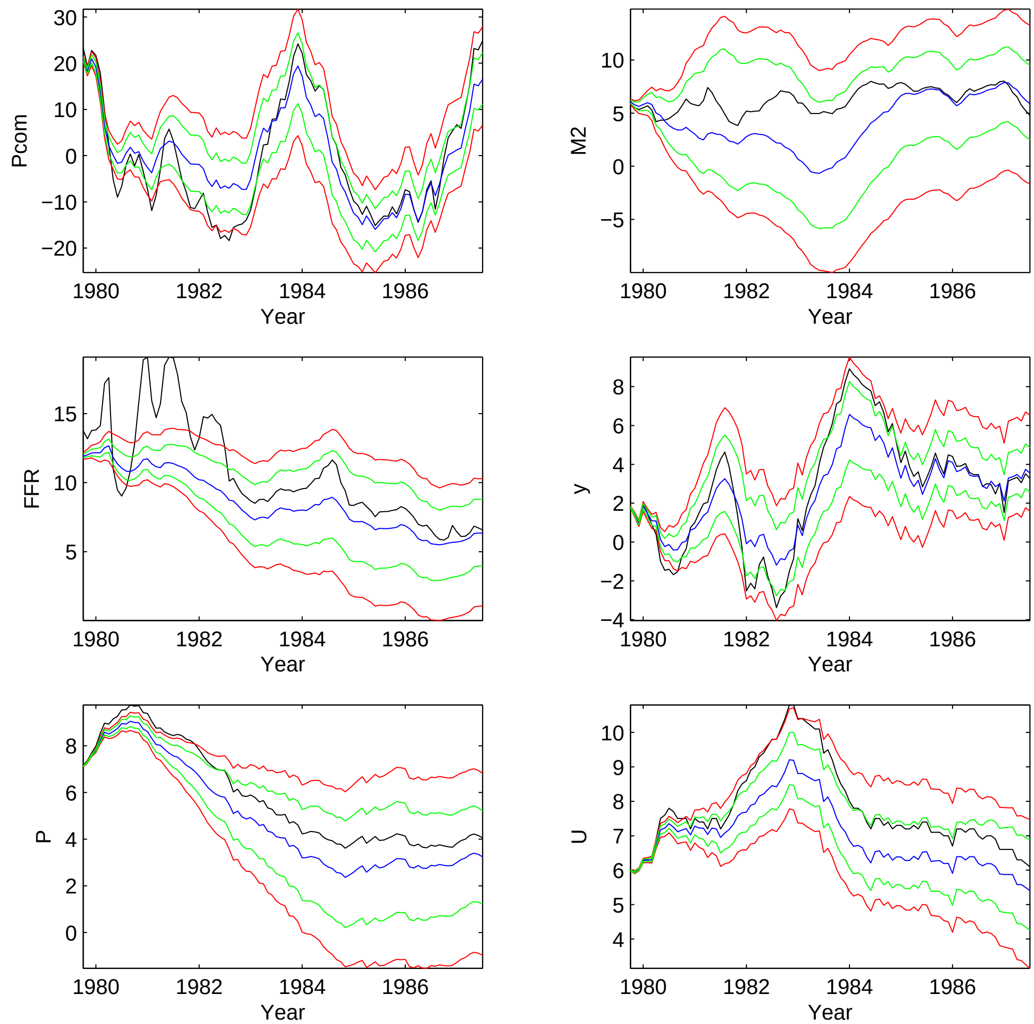

FIGURE 5. Counterfactuals: Greenspan regime through the 80's

Each graph shows the historical path (blackest), the median counterfactual path, and 68% and 90% probability bands.

as actual data with the most likely path about one percentage point lower than the historical, while the median path of unemployment would have been kept about one percentage point lower through the entire period. There would have been no tradeoff between the inflation decline and the output loss. The most likely counterfactual path of output growth would have been much smoother and the error bands imply the strong likelihood that both recessions would not have been as deep as what actually occured.

These results do not contradict some economists' view that Volcker's "harsh" stabilization policy might have been only politically, not economically, necessary. As Bryant (page 107, 1983) put it, "the policy goal of reducing inflation was pursued too zealously in 1980–82 in the United States without sufficient regard for the probable costs in unemployment and lost output."

VI.3. **Greenspan regime in place throughout 73-79.** The two rapid rises in inflation during the 70's have often been attributed to Burns's stop-go monetary policy. Indeed, our 4-state model's dynamics show no sustained regime during Burns's tenure. Our estimated monetary policy switches frequently between two regimes, with policy reacting more to money growth in one rule than the other.[12]

To quantify the effect of a different policy behavior, we re-examine the historical period 1973:1 – 1979:9 when the first upward swing of inflation was about to start at the beginning of 1973, and hold the Greenspan regime in place throughout this period. The counterfactual paths are reported in Figure 6. Since the Burns regime was actually in place for the early part of this period, replacing it by the Greenspan regime did produce different outcomes: inflation would not have been pushed as high as the first run-up in 1975, the interest rate would have been smoothed somewhat, and money growth would have been lowered in the later part of the period. The biggest difference shows up in the counterfactual paths for output growth and unemployment: a shallower recession in the early 70's with unemployment two percentage points lower in 1975, slower output growth in the later period, and a modest payoff in lower inflation at the end of the 70's.

Clearly, the inflation path with the Greenspan regime would have nonetheless resulted in a rise in inflation in the early 70's, albeit less steep than what actually occurred. The most likely counterfactual path of inflation at the end of the 70's shows no tendency to

[12]Burns (1987) acknowledged: "Partly as a result of the chronic inflation of our times, central bankers have been giving closer attention to the money supply than did their predecessors; but they continue to be seriously concerned with the behavior of interest rates." The records from the *Memoranda of Discussion* from 1970:2 to 1976:3 and FOMC transcripts in 1976–1978 suggest that the FOMC during the 70's seemed frequently to grapple with the question of whether monetary aggregates or interest rates should be used as a primary policy instrument.

drift upward but there is great uncertainty surrounding this path. Had the Greenspan policy rule been placed in the 70's, the outcome would have been better, but the differences are probably not as large as commonly thought.

VII. 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 time-varying shock variances are the most important instability in the time series of five key US macro variables. Even with the four-state model, which assumes the existence of regime changes in monetary policy, our various point estimates imply that the impact on the economy of changes in the systematic part of monetary policy are modest.

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. The truth seems to be that if there were important nonlinear elements of policy behavior in this period, thinking of them as easily detectable regime shifts is mistake.

The role of monetarism in conditioning policy responses seems to have been more important than is allowed for in most currently fashionable theories.

APPENDIX A. EXAMPLE MODEL WITH CGG IDENTIFICATION PROBLEMS

Here we display a standard simple New Keynesian model. The model has a unique equilibrium with all variables stationary.

The model is written in terms of the interest rate r, logarithmic deviation from steady state of output y, and inflation π. Its equations are

$$\text{M policy:} \qquad r_t = \alpha_0 \pi_{t-1} + \alpha_1 y_{t-1} + \alpha_2 r_{t-1} + \varepsilon_t \tag{A1}$$

$$\text{IS:} \qquad E_t y_{t+1} = y_t + \gamma(r_t - E_t[\pi_{t+1}] + \log\beta) + \xi_t \tag{A2}$$

$$\text{Phillips curve:} \qquad \pi_t = \theta_0 E_t[\pi_{t+1}] + \theta_1 E_t[y_{t+1}] + \omega_t \tag{A3}$$

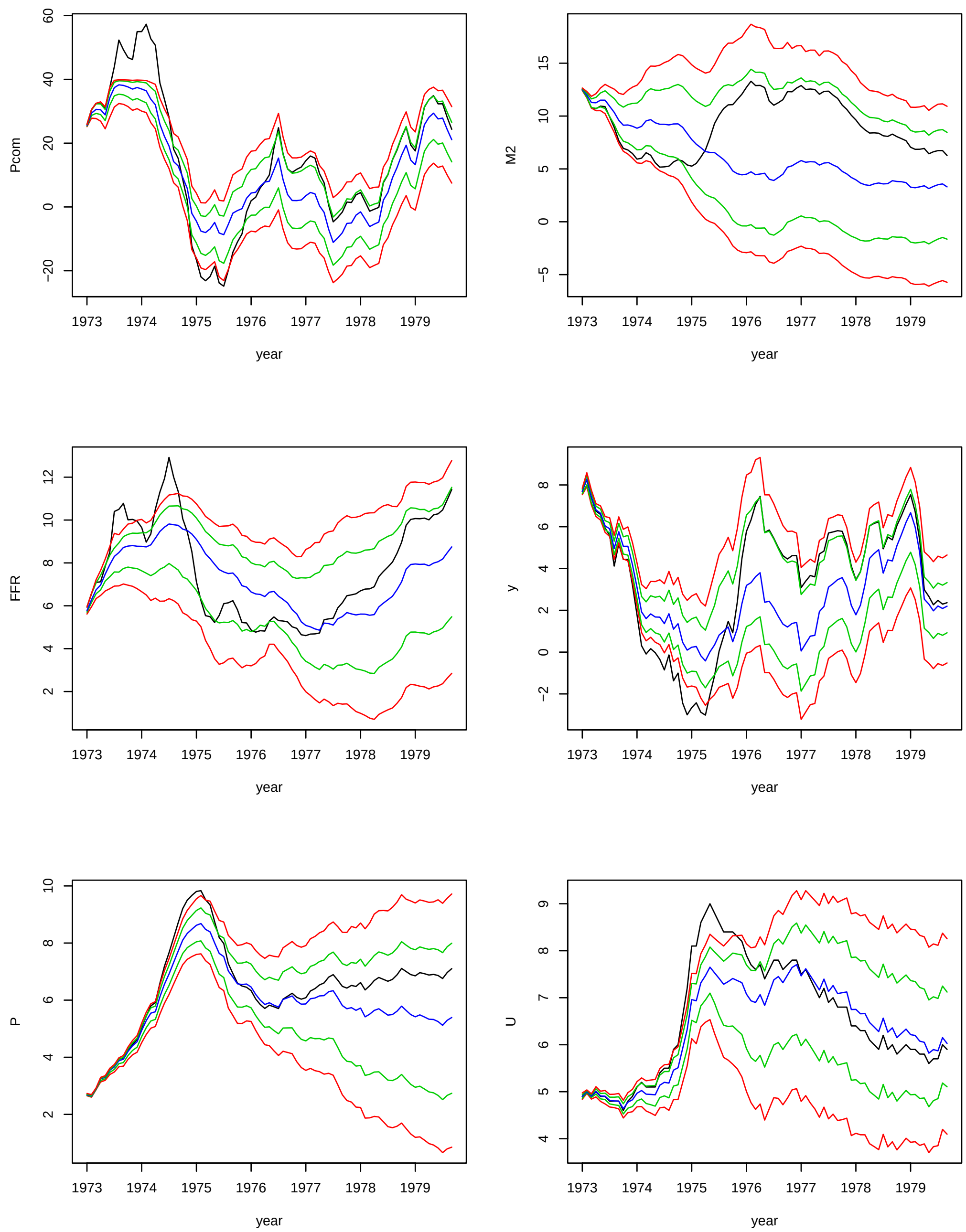

FIGURE 6. Counterfactuals: Greenspan regime through the 70's

Each graph shows the historical path (blackest), the median counterfactual path, and 68% and 90% probability bands.

(A4)

With reasonable parameter values[13], this model's solution implies that inflation is serially uncorrelated, that other variables follow MA(2) processes, and that there is a single state variable (the linear combination of lagged variables appearing on the right-hand-side of the monetary policy equation). The policy rule implies a strong long-run response of interest rates to any sustained increase in inflation (which of course does not occur in equilibrium), so there is no problem with existence or uniqueness of a solution.

Any attempt to estimate a purely forward-looking Taylor rule from data generated by this economy by instrumental variables methods would fail. Because of the one-dimensional state, there is really only one instrument available for the two expected future values on the right-hand-side of a forward-looking Taylor rule. Indeed, if twice-lagged variables were used as instruments, they would have no correlation at all with the variables they were instrumenting for. As is well known, in this weak-instrument situation, results might easily nonetheless appear to be significant.

While this result is extreme, resulting from the simplicity of the model, it illustrate problems that will be present in any model. If policy succeeds in keeping inflation low and stable, it will make variation in expected future inflation small, and may easily make high current nominal rates predict low, not high, future inflation. This is likely to make IV results erratic, as well as necessarily misleading when the Taylor rule is not in fact forward-looking.

Furthermore, if we expanded this model, say by adding more lags on the right-hand-sides of the first and third equations, so that IV methods are at least possible, they would estimate the IS equation, not the policy rule. If the second (IS) equation is renormalized to have r_t on the left, it relates current r to expected future inflation, expected future output growth, and a shock. Since this is the same form as the forward-looking Taylor rule, and the equation is distinguished from the other two by the identifying assumptions, IV methods to estimate such an equation would reproduce the IS curve, normalized on r_t as left-hand side variable. This would of course give a coefficient on expected future inflation of approximately one,

[13]For example, $\alpha_0 = .3$, $\alpha_1 = .4$, $\alpha_2 = .8$, $\gamma = 2$, $\theta_0 = .9$, $\theta_1 = .3$.

implying a high probability, given the data, of values less than one. But this would not indicate any problem with existence or uniqueness of equilibrium.

APPENDIX B. ESTIMATION AND INFERENCE

B.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 1998a 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 1998a, $\lambda_0 = 0.6, \lambda_1 = 0.1, \lambda_2 = 1.0, \lambda_3 = 1.2, \lambda_4 = 0.1, \mu_5 = 5.0$, 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 k^{th} 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.[14]

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.

B.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

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

(but restricted) parameters.

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

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

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

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

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

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 B.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 as many as 15 hours on an Intel Pentium 4 2.0GHz PC. [15]

B.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 follow Sims and Zha 2004 to use a Gibbs sampler to obtain the joint distribution $\pi(\theta, S_T \mid Y_T)$

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

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 from (for details, see Sims and Zha 2004).

B.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.[16] 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

[16]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.

probabilities of each state for all $t \in \{1, \ldots, T\}$ match closest to the posterior estimates of those probabilities.[17]

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.

REFERENCES

BERNANKE, B., AND I. MIHOV (1998): "Measuring Monetary Policy," *Quarterly Journal of Economics*, 113(3), 869–902.

BOIVIN, J. (2004): "Has U.S. Monetary Policy Changed? Evidence from Drifting Coe cients and Real-Time Data," Discussion paper, Columbia University.

BOIVIN, J., AND M. GIANNONI (2003): "Has Monetary Policy Become More Effective?] Jean Boivin Marc Giannoni Working Paper 9459," Working Paper 9459, National Bureau of Economic Research.

BRYANT, R. C. (1983): *Controlling Money: The Federal Reserve and Its Critics*. The Brookings Institute, Washington, D.C.

BURNS, A. F. (1987): "The Anguish of Central Banking," *Federal Reserve Bulletin*, September, 687–698.

CHAPPELL., H. W., AND R. R. MCGREGOR (2000): "Fed Chat: FOMC Transcripts and the Politics of Monetary Policymaking," Manuscript, University of South Carolina and University of North Carolina at Charlotte.

CHIB, S. (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.

[17]This label normalization is a computationally efficient way to approximate Wald normalization discussed by Hamilton, Waggoner, and Zha 2003.

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 (2002): "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S..," Discussion paper, Arizona State University and New York University, `http://www.stanford.edu/~sargent/`.

COOLEY, T. F., S. F. LEROY, AND N. RAYMON (1984): "Econometric policy evaluation: Note," *The American Economic Review*, 74, 467–470.

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

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

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

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

HAMILTON, J. D., D. F. WAGGONER, AND T. ZHA (2003): "Normalization in Econometrics," Manuscript, University of California (San Diego) and Federal Reserve Bank of Atlanta.

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

LEEPER, E. M., C. A. SIMS, AND T. ZHA (1996): "What Does Monetary Policy Do?," *Brookings Papers on Economic Activity*, 2, 1 – 78.

LEEPER, E. M., AND T. ZHA (2003): "Modest Policy Interventions," *Journal of Monetary Economics*, 50(8), 1673–1700.

LUCAS JR., R. E. (1972): "Econometric Testing of the Natural Rate Hypothesis," in *The Econometrics of Price Determination*, ed. by O. Eckstein, pp. 50–59. Board of Governors of the Federal Reserve and the Social Science Research Council.

PRIMICERI, G. (2003a): "Time Varying Structural Vector Autoregressions and Monetary Policy," Discussion paper, Princeton University, `http://www.princeton.edu/~gprimice/research.html`.

——— (2003b): "Why Inflation Rose and Fell: Policymakers' Beliefs and US Postwar Stabilization Policy," Discussion paper, Princeton University, `http://www.princeton.edu/~gprimice/research.html`.

ROBERTSON, J. C., AND E. W. TALLMAN (2001): "Improving Federal-Funds Rate Forecasts in VAR Models Used for Policy Analysis," *Journal of Business and Economic Statistics*, 19(3), 324–330.

SIMS, C. A. (1987): "A Rational Expectations Framework for Short Run Policy Analysis," in *New Approaches to Monetary Economics*, ed. by W. Barnett, and K. Singleton, pp. 293–310. Cambridge University Press.

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

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

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

——— (1998b): "Does Monetary Policy Generate Recessions?," Federal Reserve Bank of Atlanta Working Paper 98-12.

——— (2004): "MCMC Method for Markov Mixture Simultaneous-Equation Models: A Note," Manuscript, Princeton University and Federal Reserve Bank of Atlanta.

WAGGONER, D. F., AND T. ZHA (2003): "Likelihood Preserving Normalization in Multiple Equation Models," *Journal of Econometrics*, 114(2), 329–347.

DEPARTMENT OF ECONOMICS, PRINCETON UNIVERSITY,FEDERAL RESERVE BANK OF ATLANTA
E-mail address: `sims@princeton.edu, tzha@mindspring.com`