DRIFT AND BREAKS IN MONETARY POLICY

CHRISTOPHER A. SIMS

ABSTRACT. US monetary policy behavior since 1948 is modeled as nonlinear, changing over time according to a hidden Markov chain specification. Though the estimated Markov chain model implies large shifts in the form of the policy reaction function, its improvement in fit over a simple linear model comes almost entirely from its allowance for persistent heteroscedasticity. A linear model that allows such heteroscedasticity fits almost as well. The shifts in policy regime that are uncovered are not unidirectional—they oscillate, with a given state seldom persisting more than a few years. The paper discusses how these results mesh with attempts to interpret this period through the Lucas critique and natural rate models.

I. INTRODUCTION

The centerpiece of this paper is a stylized model of the behavior of US monetary policy authorities over 1948-97. It is therefore a contribution to the "reaction function" literature. Unlike most of that literature, however, it considers skeptically the notion that this period can be broken into disjoint periods of distinctly different monetary policy behavior. It insists on modeling possible deviation from a simple linear reaction function with an explicit stochastic model of the deviations. It does so using the hidden Markov chain approach that has recently been applied by Chib (1996) and Kim and Nelson (1999) and earlier Hamilton (1989).

The paper's substantive econometric conclusions are

- there is strong evidence of deviation from a linear, Gaussian policy reaction function;
- there are variations over time in the size of disturbances to the policy rule that are more important to model fit than the variation in the coefficients of the reaction function, with the two types of variation occurring together;

Date: July 2, 1999.

Key words and phrases. change point, Bayesian methods, vector autoregression, identification, time-varying parameters, Lucas critique.

- because of the model's logarithmic form, all of the three estimated policy regimes involves strong enough eventual reaction of the interest rate to the price level to guarantee existence and uniqueness of equilibrium under a passive[1] fiscal policy. In the neighborhood of historically observed inflation rates, though, the linear approximation to only one of the three rules reacts strongly enough to guarantee a unique price level.
- the variation in policy is not in the form of unidirectional flow from one regime to another, as would be expected if policy changes reflected steady improvement based on learning; it is in the form of repeated oscillations among types of behavior, with "regimes" generally lasting well under a decade;

The results suggest that there has by and large been continuity in American monetary policy, albeit with alternation between periods of erratic, aggressive reaction to the state of the economy and periods of more predictable and less aggressive response. The aggressive type of policy seems to be more associated with periods of flat or declining interest rates than with rising interest rates. Not only is there no evidence of the kind of one-time change in policy rule, accompanied by changes in the cyclical behavior of real variables, that is the workhorse of theoretical rational expectations policy analysis, there is little evidence that monetary policy makers at the beginning of the period had any tendency to be less "tight".[2]

That the nonlinearities in policy behavior we have found do not fit the categories of standard rational expectations theory does not mean they are uninteresting. It should be possible to use hidden-state policy models like this to probe for nonlinear impacts of monetary policy on the economy. The predictability of monetary policy could in principle be important. Do the strong variations in predictability found here have effects on the behavior of the economy? This seems an important topic for further research, and this paper includes discussion of how it could be pursued.

II. Some Descriptive Statistics

Discussion of the recent history of US monetary policy by the economic punditry sometimes assumes that it began the postwar era in the thrall of Keynesianism, and was thus not vigilant about controlling inflation. The result of this policy is then taken to have been the rise of inflation in the 60's and 70's, which was cut short, it is thought, by the advent of a new monetary policy that recognized the need to

[1]This is the terminology of Leeper (1991). A very similar classification of fiscal policies by Woodford (1996) labels this type of policy "Ricardian".

[2]The data, seen through this model, thus fail to confirm either of the competing stories about how American inflation ended that Sargent (1999) discusses.

control inflation and the absence of any permanent tradeoff between inflation and unemployment.

Even a quick examination of the facts shows that this story is hard to see in the data. We see in Figure 1 a plot of a short interest rate against the consumer price index (CPI) inflation rate for this period. Clearly nominal interest rates have tracked inflation rates closely throughout the period. For the inflation rate to be determinate and stable, monetary policy must eventually increase interest rates at least in proportion to sustained increases in the inflation rate.[3] From Figure 1 we see that this criterion appears to have been met throughout the period. However, in the 70's, as energy price shocks created sudden inflation, the real rate was allowed to stay at approximately zero over a prolonged period. In hindsight, knowing that this inflation burst was not a temporary burst like the Korean war inflation of the early 50's or the temporary "cost push" inflation of the '57-58 recession, it can be argued that pushing the interest rate up above the inflation rate would have been advisable. But note that even the sharp rise in the interest rate up to its peak at the beginning of 1980 merely tracked the behavior of CPI inflation. It was only after inflation peaked and began to decline that the interest rate, which followed downward with a delay, opened up a gap implying a real rate at the level (or slightly above the level) that had prevailed in the 60's.

One can see from the graph that the response of interest rates to inflation was probably slower, though perhaps not much less strong in total, in the 60's, that the level of interest rates relative to inflation was maintained higher then, and that interest rates followed a smoother time path month to month in the 60's than in the 70's and 80's. But it remains to be seen whether these differences across periods are sharp enough to give nonlinear models a clear edge in fit.

III. The Model

Because of this paper's exploratory nature, the model of policy behavior has been kept drastically simple. We model monetary policy as setting a short term interest rate (the 6 month commercial paper rate) as a function of past values of the interest rate and of a commodity price index. The aim was to use a very manageable model in which interest rates tend to rise in response to inflationary pressure and in which upward disturbances in the interest rate lead to downward movements in prices. A bivariate system in interest rates and commodity prices has just these properties, even though it of course ignores the likely dependence of monetary policy on a number of other variables.[4] We use monthly data from 1948:1-1997:10.

[3]This assumes passive fiscal policy.

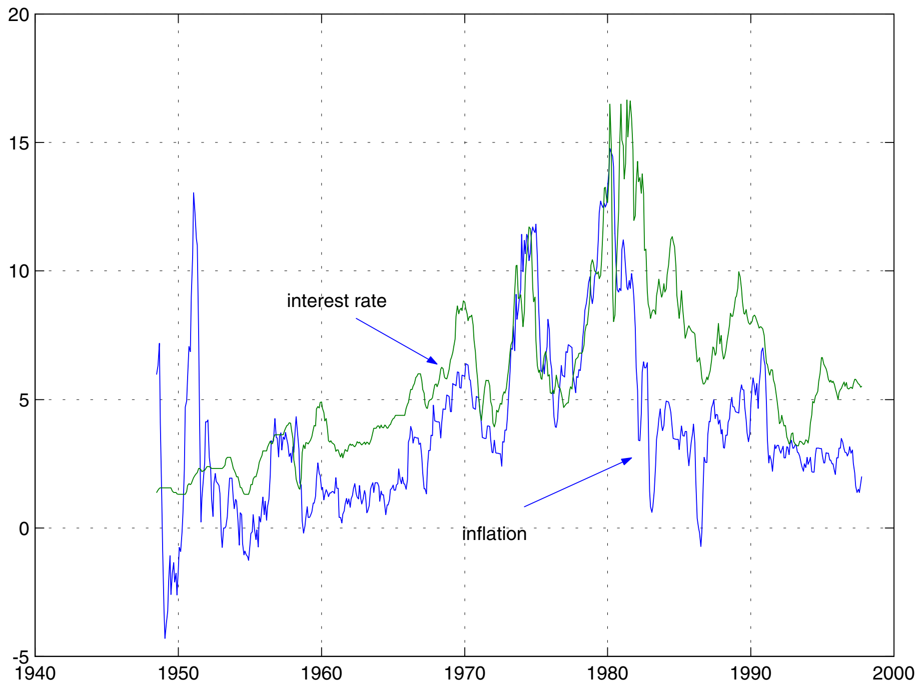

FIGURE 1. CPI Inflation and the 6-month Commercial Paper Rate

The model is described by the following equations:

$$\alpha(L)r_t = c(S_t) + b(S_t)\beta(L)p_t + \sigma(S_t)\varepsilon_t \tag{1}$$

$$P[S_t = i \mid S_{t-1} = j] = p_{ij}\ ,\ i = 1, \ldots, 3\ ,\ j = 1, \ldots, 3\ , \tag{2}$$

where ε_t is i.i.d. $N(0,1)$ conditional on the entire $\{S_t\}$ sequence and on $\{r_s, p_s, s < t\}$. We normalize the initial coefficients of the lag polynomials, α_0 and β_1, to 1 (setting $\beta_0 = 0$) and make them 6th order. In this equation r is the natural log of the 6-month commercial paper rate, expressed as a percent at annual rates, and p is the log of a commodity price index.

The transition matrix $P = [p_{ij}]$ for the unobserved discrete Markov process S is controlled by three parameters π_1, π_2, and π_3, so that it has the form

$$P = \begin{bmatrix} \pi_1 & 1-\pi_1 & 0 \\ (1-\pi_2)/2 & \pi_2 & (1-\pi_2)/2 \\ 0 & 1-\pi_3 & \pi_3 \end{bmatrix} . \tag{3}$$

[4]It is standard to use the Federal Funds rate in reaction function research. The commercial paper rate has the advantage that it will allow extension of the study to the interwar years, though nothing is done in this draft with interwar data.

The model implies that the policy reaction function varies among three distinct forms as S_t varies among its three possible values 1 to 3. The coefficients on lagged r are the same in all three regimes, as is the form of the lag distribution on p. However, the scale of the lag distribution on p varies across regimes, as does the regression's constant term and its disturbance variance.

While this model is nonlinear and mixes discrete and continuous components of random variation, it falls in the class of what Chib (1996) calls hidden Markov chain models. This means that its likelihood, the pdf of the data conditional on parameters (but integrated over S) can be calculated recursively analytically, and also that draws from its posterior pdf can be generated easily. The model in Hamilton (1989) is also in this class, and he exploited its recursive tractability. A variety of models of this class are discussed in Kim and Nelson (1999).

There is clearly a lot of room for making this model more complicated. We could allow the coefficients of α and β to vary with S. We could add more right-hand-side variables, with or without S-dependence in their coefficients. There is no magic in the assumption of 3 states—we could allow more. And of course a completely general specification for the P matrix would have six free parameters instead of three.

The form of the P matrix is important to interpretation, and we will discuss below the effect of varying it. We also display the effects of reducing (but not of increasing) the number of states. But note that we already have here a system with 23 free parameters. An unconstrained 4 variable, 6 lag VAR with a constant term has just 26 free parameters per equation (counting the disturbance variance), and we are well used to the idea that such a system is likely to forecast poorly unless it is estimated with a proper prior distribution. Since this paper is to some extent breaking new ground, we maintain the 23-parameter limit for now. Complications may well be worth examining, probably with proper priors, in the future.

Details of how the likelihood is formulated and estimated are described in the appendix.

IV. Estimates of the Reaction Function Model

Maximum likelihood estimates of the parameters of the model, with associated asymptotic approximations to the standard deviations of the distribution of the true parameter values about the estimates,[5] are presented in the left half of Table 1. Note

[5]Here we interpret minus the inverse second derivative matrix of the log likelihood as the covariance matrix of a Normal approximation to the likelihood function. Such an approximation will become accurate in large samples under broad regularity conditions. Unlike the usual Normal asymptotic approximation to the distribution of the coefficient estimates about the true value, this approximate characterization of the shape of the likelihood is valid even in the presence of unit-root nonstationarity.

that the three parameters that depend on the state, c, b and σ, all vary monotonically with the state. So state 1 has low response to inflation and low disturbance variance, state 3 has high response to inflation and high disturbance variance, and state 2 is intermediate. The fact that state 3 implies the strongest response of interest rates to inflation does not imply that it is strictly "tighter" than the other two regimes. The constant terms decrease as we go from state 1 to state 3. It turns out that with inflation constant, state 1 implies a higher interest rate than state 2 at annual inflation rates below 60%, state 1 implies a higher interest rate than state 3 at inflation rates below 16%, and state 2 implies a higher inflation rate than state 3 at inflation rates below 14%. Thus the lower-numbered states imply *higher* interest rates that are less responsive to inflation, except when commodity price inflation rates reach the levels observed only at the peak of postwar inflation or in brief bursts at other times in the period.

Responses to inflation, as measured by b, vary by a factor of 8 between states 1 and 3, and standard deviations of disturbances vary by a factor of 4. The usual asymptotic $\chi^2(2)$ statistic for equality is 174 for the variance parameters and 10.5 for the b parameters. Notice that though the difference in size of the parameters is bigger for b, the standard errors for b are relatively bigger. In fact, by the Schwarz criterion the differences among states in b are (just barely) not sufficient to favor this model allowing variation in b across states over a model that restricts b to be constant across states. The asymptotic arguments that justify the Schwarz criterion may not apply in a model like this one that implies high likelihood for unit roots in the dynamics. However, direct application of the method of Laplace[6] (from which the Schwarz criterion is derived) would in this instance even more strongly favor the simpler model.

As a check on the validity of the asymptotics[7], we re-estimate the model with variation across states in the disturbance variances only, obtaining the results shown in the right half of Table 1. This produces an asymptotic $\chi^2(4)$ statistic of 10.18, which

[6](Schervish 1995, section 7.4.3). The method of Laplace is simply the idea of approximating posteriors, likelihoods, or products of these objects with functions of the underlying random variable, by second order Taylor expansions of their logs about their peaks—in other words, by their usual asymptotic Normal estimators. The Schwarz criterion follows from this approach under the assumption that the second derivative of the log likelihood is growing as $O(T)$, where T is sample size. In other contexts, where this assumption is not valid but the asymptotic normal approximation still holds (as in the presence of unit roots), the role of $\log T$ in the Schwarz criterion is taken over by the difference in log determinants of the second derivative matrices for the restricted and unrestricted models.

[7]We can check the validity of the asymptotics because we are using Bayesian asymptotics, which only make assertions about the shape of the likelihood, which can be examined directly.

	Full Model		Variances Only	
parameter	estimate	std. error	estimate	std. error
c	0.0091	0.0079	0.02204	0.094115
	0.0023	0.0149		
	-0.0422	0.1143		
b	0.2869	0.0596	0.2476	0.84093
	0.3559	0.1559		
	2.5163	0.6908		
α	1.4320	0.0429	1.4424	0.16037
	-0.5925	0.0703	-0.60474	0.32806
	0.2668	0.0681	0.27231	0.30502
	-0.1386	0.0662	-0.15331	0.084267
	0.0304	0.0646	0.046084	0.077247
	-0.0063	0.0369	-0.010583	0.061189
β	-0.6170	0.0101	-0.1960	3.0638
	-0.7078	0.2098	-1.3872	4.9638
	0.8687	0.2975	1.3887	2.002
	-0.4225	0.1388	-0.5951	5.0373
	-0.1164	0.1432	-0.2165	5.2359
σ	0.0259	0.0015	0.0258	0.0050
	0.0530	0.0029	0.0529	0.0203
	0.1110	0.0127	0.1221	0.0149
π	0.9748	0.0093	0.9764	0.0183
	0.9578	0.0152	0.9591	0.0248
	0.9093	0.0336	0.9192	0.0105
log LH	1005.75		1000.66	

Table 1. MLE parameter estimates of preferred model

Note: The parameters c, b, and σ have three values each, listed in order for the states we label 1, 2, and 3. The coefficients in α are those on lags 1 through 6 of r on the right-hand side of the equation. The coefficients in β are the scaled coefficients on lags 2 through 6 of p on the right-hand side, with the scaled coefficient on the first lag normalized to one and b providing the scale factor. See equation (1) in the text. The listed standard errors are based on the square roots of the diagonal elements of minus the inverse of the second derivative of the log likelihood, as approximated by the BFGS update process during calculation of the MLE.

is at a marginal significance level of .038. In a sample of size 592, this is weak evidence in favor of the null, since the Schwarz criterion would require that the statistic be 25.5 before it favored the larger model. Furthermore, note that in the restricted model the parameters b and β are locally very ill-determined. Direct application of the method of Laplace would therefore imply even stronger weight toward the more restricted model, because in effect the restricted model has nearly redundant parameters, and is therefore effectively smaller than a count of its parameters would suggest.

It might appear from the "insignificance" of b and β estimates in the restricted model that the evidence for any effect of p on r at all is weak. This is not true, however. When the model is restricted to allow no effect of p on r, so that it is just a univariate model of r with switching among residual variance regimes and constant terms, the log likelihood maximum is reduced to 983.15. Eliminating the 6 parameters associated with p therefore produces a $\chi^2(6)$ statistic of $2 \times (1005.75 - 983.15) = 43.2$, which has a marginal significance of around 10^{-7} and favors the model including prices by the Schwarz criterion.

It is difficult to assess the strength of the estimated response of r to p from these individual coefficients. In the full model, one minus the sum of coefficients on lagged r is 0.00812269, with a standard error of .002624. So even though the sum is small, implying a root close to one in $\alpha(L)$, likelihood sharply distinguishes the sum from zero. The sum of coefficients in $\beta(L)$ is .0050 with a standard error of .021, so the estimated long run response of the level of interest rates to the level of commodity prices is not distinguished from zero, by model fit. The point estimate of the long run response of the log of the interest rate to the log of commodity prices is about $.62b(S_t)$, which ranges from .18 to 1.56. The largest of these implies that to invoke an increase in the interest rate from, say, 5% to 6%, the price level would have to increase by 13%, so the estimated responses are quite modest.

The cumulative sum of coefficient in $\beta(L)$ is the lag distribution on monthly inflation rates. If we take b times the first 5 elements of this cumulative sum as the lag distribution on inflation (setting to zero the 6th and later coefficients, which are of course all .0050), sum the result, and divide by $\alpha(1)$, we have an estimate of the long run response of $\log r$ to monthly inflation, under the assumption of zero long run response to the price level. Dividing the result by 12 to get responses to annual (rather than monthly) inflation rates gives us, as long run responses in the three states, 5.09, 6.31 and 44.6. Because of the logarithmic specification, the interest rate levels above which these estimates imply more than one-for-one response of r to the commodity price inflation rate are 20%, 16%, and 2.2%, respectively. It is interesting to note that the middle state implies more than one-for-one response starting at just the level of interest rates actually reached in the late 70's and early 80's

Because the logarithmic specification matters for the interpretation of the model, it is worth noting that the likelihood, corrected for the Jacobian of the data transformation, is quite a bit lower when the model is estimated in the levels of the interest rate.

The model was also estimated with symmetric jumping among the states, i.e. with $p_{ij} = (1 - \pi_i)/2$, $j \neq i$. This results in a slightly lower log likelihood of around 1003. The MLE for this symmetric model implies that only two of 16 likely points of transition between states in the sample involved jumps between non-adjacent states, an unlikely result if jump probabilities are in fact symmetric. The model was also estimated with the states restricted to form an irreversible chain, i.e. with $p_{12} = 1-\pi_1$, $p_{23} = 1 - \pi_2$, $p_{ij} = 0$ for all other $j \neq i$, and $\pi_3 = 1$. This specification has one fewer free parameter, and produces much lower log likelihood (around 960).[8]

The post-sample probabilities of state 1 and state 3 (that for state 2 of course is 1 minus the sum of the other two), together with the dependent variable in the regression, are plotted in Figure 2. Note that the high-variance, high-response state 3 is estimated as likely only during the early 80's reserve-targeting period, plus a brief period in the 50's when the interest rate came down sharply and a similar but even briefer period in the mid-70's. The rest of the period shows regular oscillations between states 1 and 2, except for the 60's, during which there was an unusually long period of prevalence of state 1. Futhermore, the advent of state 3 in every case coincides with the end of a period of rising interest rates. It seems to be required to explain the rapid declines in rates that occur at the starts of recessions and the ends of inflations. The switch to state 3 at the start of the 80's occurs in the first month of 1980, just before the interest rate peak of 16.5% the next month, which was then only very slightly exceeded in two later months that year. Most of the period 1980:1-1982:10 during which this state is estimated to have persisted is a time of constant or declining interest rates and declining inflation. The responsiveness of interest rates to inflation in this state explains the speed with which interest rates declined over this period, though it also contributes at the beginning of the period, when lagged inflation was high, to the high level of interest rates.

V. Implications of the Estimates

The case for recognizing time variation in the variance of disturbances in the reaction function is very strong. Despite the unrealistically simplified form of this model, there seems little doubt that this result would carry over to more realistic specifications. That there is heteroscedasticity in policy reaction function disturbances in

[8] The log likelihoods cited here are vague, because the exact numerical results were lost as this draft is being written, leaving no time to reproduce them.

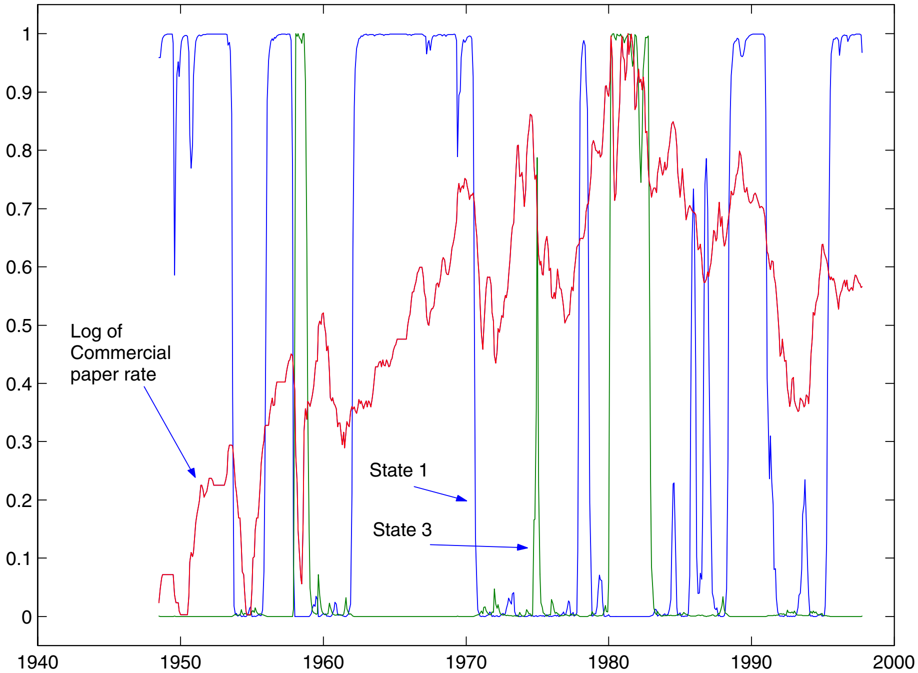

FIGURE 2. Post-Sample Probabilities of States

linear VAR's has has been recognized before. This paper shows that it is well modeled as persistent.[9]

The fact that the states seem to arrange themselves one-dimensionally suggests that it might make sense to expand to more than three, while keeping the parameterization from expanding too much by constraining P to take a form that concentrates along its diagonal. There would be no computational problem in such an expansion if it could keep the P matrix dependent on no more than 3 or 4 parameters. One way to do this would be to choose P so that it can be thought of as a discrete approximation to a continuous AR process.

One should treat more cautiously substantive conclusions about the nature of policy based on this model, because they could easily shift with the addition of more variables or the use of the non-recursive identification schemes that we know can give better

[9]The symmetric jumping form of the model includes as a special case equal probabilities on all states in all periods, which would imply a simple mixture of Gaussian disturbance distributions, i.e. fat tails rather than persistent heteroscedasticity. The point estimates of that model were very far from this form. However, we have not directly checked for a general three-state mixture model without persistence, which would be what is implied by having all columns of P constant, but not necessarily at the same level.

results.[10] Nonetheless, this model is no more oversimplified than some that have been used to spin interesting stories about policy, so it is interesting to proceed.

The results present a paradox, in that they show that a story that asserts a sudden, large increase in the strength of reaction to inflation in the early 80's is consistent with the data, yet that the improvement in fit from modeling this change in the reaction to inflation is quite modest. My own interim conclusion is that it is not yet time to toss out linear reaction function models entirely, so long as they are amended to allow persistent heteroscedasticity.

The point estimates of the full model, though consistent with a toughening of anti-inflationary policy in the early 80's, do not imply an irreversible learning process. Indeed fit deteriorates badly if policy is not allowed to revert to previous forms. The 80's are estimated as quite different, in other words, but also as quite temporary. One interpretation of this pattern fits the pessimistic conclusion of Sargent (1999): the experience of the seventies taught policy-makers the importance of anti-inflationary action, but the accumulating experience of a stable economy in the late 80's and 90's could be simply undoing the learning process, preparing the stage for another inflationary crisis.

My own preferred interpretation would treat the policy rule as stable, though non-linear. Policy reacts to commodity price inflation, but it does not react strongly to temporary fluctuations in it, which are frequent. It reacts with increasing strength as the evidence of strong and persistent inflation accumulates, and is capable of very vigorous action when the threat of inflation is very clear. If the public understands this pattern of behavior, theory implies that the episodes in which strong counter-inflationary action is needed may be rare. Indeed it is easy to construct models in which, so long as the public is convinced of the readiness of the policy authority to react strongly, the conditions under which the policy authority needs to do so never arise.

The most interesting next directions for research, it seems to me, are checking for whether the states that are useful in modeling nonlinearity in policy behavior also have direct effects on the private sector. That we should look for this possibility is the real lesson of the Lucas critique. Policy changes need not be well modeled as zero-mean, stationary random shifts in the constant term of a linear reaction function. They could instead need to be modeled as stochastic, more or less persistent, changes in the coefficients or variance parameters of those relationships. If they do have this character, extrapolating the effects of policy changes from linear approximations to

[10]Much of the identified VAR literature using postwar data has stuck to recursive identification schemes, probably as a matter of convenience. Sims (1999) shows that extending structural VAR methods to the interwar period seems to require use of non-recursive identification.

the true policy rule could be quite misleading.[11] To take account of the Lucas critique in practice, we need to look for nonlinearity in policy behavior, and then to examine the connections between the sources of variation in policy behavior and their effects (if any) on private sector behavior. To do this, we must identify variation that originates in policy behavior and distinguish it form variation originating in the private sector, a difficult task. As the structural VAR literature has shown, however, it is possible to approach this task without insisting that the model be formulated entirely in terms of fully interpretable "deep" parameters of taste and technology.

It is quite possible that the nonlinearities in policy behavior that this paper's model points to are unimportant to modeling the reaction of the public to monetary policy. That is, the public's behavior might depend on the history of interest rates and reserves directly and more or less linearly, without regard to the current state of the policy reaction function. But rational expectations has given us a rich set of examples to show that it could be that the public reacts differently to the same pattern of changes in interest rates under policy states 1 and 3. Further exploration, in multivariate models, to see whether this possibility emerges as practically important, is clearly desirable. It is likely to have the additional benefit that, by bringing more information to bear, it should sharpen inference on the fundamental question that this paper's results leave open: In telling interesting stories about policy shifts with this papers's Markov chain models, are we just making a man in the moon out of random meteor craters?

Appendix A. Calculating the Likelihood

Following Hamilton (1989), we treat the initial lagged values of r_{1-s} and p_{1-s}, $s = 1, \dots, 6$, as independent of S_1, treat S_1 as drawn from its steady-state distribution (as implied by P), and condition on the initial lagged values of r and p in forming the likelihood.

We will use $\mathcal{I}_t$ to refer to the observable information set at t, i.e. $set r_{t-s}$, p_{t-s}, $s \geq 0$. The model implies

$$q(r_t \mid \mathcal{I}_t, S_t) = \frac{1}{\sigma(S_t)} \phi\left(\frac{\alpha(L)r_t - b(S_t)\beta(L)p_t - a(S_t)}{\sigma(S_t)}\right). \tag{4}$$

Here as elsewhere we use q as a generic symbol for a pdf when the random variable and the conditioning set are apparent from q's arguments. Also, we use ϕ as the symbol for a standard $N(0, 1)$ Gaussian pdf. We can form the joint one-step-ahead

[11] This way of looking at the Lucas critique is not new. See Cooley, Leroy, and Raymon (1984) andSims (1987).

density

$$q(r_t, S_t \mid \mathcal{I}_{t-1}) = q(S_t \mid \mathcal{I}_{t-1}) q(r_t \mid \mathcal{I}_{t-1}, S_t) \tag{5}$$

and from this the one-step-ahead predictive density for r_t as

$$q(r_t \mid \mathcal{I}_{t-1}) = \sum_{S_t=1}^{3} q(r_t, S_t \mid \mathcal{I}_{t-1}) \,. \tag{6}$$

We can form

$$q(S_t \mid \mathcal{I}_t) = \frac{q(r_t, S_t \mid \mathcal{I}_{t-1})}{\sum_{S_t} q(r_t, S_t \mid \mathcal{I}_{t-1})} \,. \tag{7}$$

This last step rests on the assumption that the distribution of $p_t \mid \mathcal{I}_{-1}, r_t, S_t$ is independent of S_t. In other words, all influence of policy on current p comes through current r, with no direct perception of and reaction to S_t by the public within the current period.

To complete the recursion, then, we form

$$q(S_{t+1} \mid \mathcal{I}_t) = \begin{bmatrix} q(S_t = 1 \mid \mathcal{I}_t) \\ q(S_t = 2 \mid \mathcal{I}_t) \\ q(S_t = 3 \mid \mathcal{I}_t \end{bmatrix}' \cdot P \,. \tag{8}$$

With (8) in hand, we can return to (4) and (5) to generate the time-$t+1$ predictive pdf, and thus recursively the entire likelihood.[12] To start off the algorithm, we need only have in hand $q(S_0 \mid \mathcal{I}_0)$, which as we stated at the start is taken to be given by the steady-state distribution of S, extracted as a left eigenvector of P corresponding to an eigenvalue of 1.

With this recursion complete for the whole sample $t = 1, \ldots, T$, we can then execute a backward recursion to evaluate $q(S_t \mid \mathcal{I}_T)$. We can do this because

$$q(S_{t+1}, S_t \mid \mathcal{I}_T) = \kappa(\mathcal{I}_T) q(S_{t+1} \mid \mathcal{I}_T) q(S_{t+1} \mid S_t) q(S_t \mid \mathcal{I}_t) \,, \tag{9}$$

where $\kappa(\mathcal{I}_T)$ is a normalizing constant set so that the expression sums to one across S_t, S_{t+1}. We can therefore find the desired $q(S_t \mid \mathcal{I}_T)$ by summing (9) over S_{t+1}. That (9) holds depends on two assumptions we have made:

$$q(S_{t+1} \mid S_t, \mathcal{I}_t) = q(S_{t+1} \mid S_t) \tag{10}$$

$$q(r_t, p_t \mid \mathcal{I}_t, \{S_v,\, v \leq t\}) = q(r_t, p_t \mid \mathcal{I}_t,\, S_t) \,. \tag{11}$$

[12] Of course in models like this with predetermined, but possibly not strictly exogenous, right-hand side variables, what we are forming is not a complete pdf for the observed data given parameters or even a conditional pdf for it. It is a factor of the complete pdf that includes all the components that depend on the parameters we are interested in; so though it is not a pdf for the data, it is still proportional to the likelihood function.

The class of models we consider here does not include the models labeled "Markov switching models" in most of the literature. Those models, following the specification in Hamilton (1989), take the form

$$\alpha(L)(y_t - \mu_{S_t}) = \varepsilon_t \, . \tag{12}$$

Even if S in this model is a first-order Markov process, the appearance of higher than a first order lag term in α will require that the full model for y and S treat the state vector as including as many lagged values of S as of y. This is not as much of a computational problem as it might sound. The point here is not that Hamilton-style Markov switching models are any worse than the ones that have been fitted in most of this paper, but that they are different.

In the computations, to avoid the need to handle constraints on the parameters explicitly, the $\sigma(S_t)$ parameters were logged and the π parameters were given a logit transformation (i.e., $\log(\pi_t/(1-\pi_t))$). The reported parameters in Table 1 undo these transformations for both the parameter values and their estimated standard errors.

References

Chib, S. (1996): "Calculating Posterior Distributions and Modal Estimates in Markov Mixture Models," *Journal of Econometrics*, 75, 79–98.

Cooley, T. F., S. F. LeRoy, and N. Raymon (1984): "Econometric Policy Evaluation: Note," *American Economic Review*, 74, 467–70.

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

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

Leeper, E. M. (1991): "Equilibria Under Active and Passive Monetary And Fiscal Policies," *Journal of Monetary Economics*, 27, 129–47.

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

Schervish, M. J. (1995): *Theory of Statistics*, Springer Series in Statistics. Springer, New York.

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.

——— (1999): "The Role of Interest Rate Policy in the Generation and Propagation of Business Cycles: What Has Changed Since the 30s?," in *Proceedings of the 1998 Annual Research Conference*. Federal Reserve Bank of Boston.

Woodford, M. (1996): "Control of the Public Debt: A Requirement for Price Stability?," Working Paper 5684, NBER.