

# Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S.\*

Timothy Cogley  
Arizona State University

Thomas J. Sargent  
New York University and Hoover Institution

Revised: October 2002

## Abstract

For a VAR with drifting coefficients and stochastic volatilities, we present posterior densities for several objects that are of interest for designing and evaluating monetary policy. These include measures of inflation persistence, the natural rate of unemployment, a core rate of inflation, and ‘activism coefficients’ for monetary policy rules. Our posteriors imply substantial variation of all of these objects for post WWII U.S. data. After adjusting for changes in volatility, persistence of inflation increases during the 1970s then falls in the 1980s and 1990s. Innovation variances change systematically, being substantially larger in the late 1970s than during other times. Measures of uncertainty about core inflation and the degree of persistence covary positively. We use our posterior distributions to evaluate the power of several tests that have been used to test the null of time-invariance of autoregressive coefficients of VARs against the alternative of time-varying coefficients. Except for one test, we find that those tests have low power against the form of time variation captured by our model. That one test also rejects time invariance in the data.

## 1 Introduction

This paper extends the model of Cogley and Sargent (2001) to incorporate stochastic volatility and then reestimates it for post World War II U.S. data in order to shed light on the following questions. Have aggregate time series responded via time-invariant linear impulse response functions to possibly heteroskedastic shocks? Or

---

\*We are grateful to Jean Boivin, Mark Gertler, Sergei Morozov, Simon Potter, Christopher Sims, and Tao Zha for comments and suggestions.

is it more likely that the impulse responses to shocks themselves have evolved over time because of drifting coefficients or other nonlinearities? We present evidence that shock variances evolved systematically over time, but that so did the autoregressive coefficients of VARs. One of our main conclusions is that much of our earlier evidence for drifting coefficients survives after we take stochastic volatility into account. We use our evidence about drift and stochastic volatility to infer that monetary policy rules have changed and that the persistence of inflation itself has drifted over time.

## 1.1 Time invariance versus drift

The statistical tests of Sims (1980, 1999) and Bernanke and Mihov (1998a, 1998b) seem to affirm a model that contradicts our findings. They failed to reject the hypothesis of time-invariance in the coefficients of VARs for periods and variables like ours. To shed light on whether our results are inconsistent with theirs, we examine the performance of various tests that have been used to detect deviations from time invariance. Except for one, we find that those tests have low power against our particular model of drifting coefficients. And that one test actually rejects time invariance in the data. These results about power help reconcile our findings with those of Sims and Bernanke and Mihov.

## 1.2 Bad policy or bad luck?

This paper accumulates evidence inside an atheoretical statistical model.<sup>1</sup> But we use the patterns of time variation that our statistical model detects to shed light on some important substantive and theoretical questions about post WWII U.S. monetary policy. These revolve around whether it was bad monetary policy or bad luck that made inflation-unemployment outcomes worse in the 1970s than before or after. The view of DeLong (1997) and Romer and Romer (2002), which they support by stringing together interesting anecdotes and selections from government reports, asserts that it was bad policy. Their story is that during the 1950s and early 1960s, the Fed basically understood the correct model (which in their view incorporates the natural rate theory that asserts that there is no exploitable trade off between inflation and unemployment); that Fed policy makers in the late 1960s and early 1970s were seduced by Samuelson and Solow's (1960) promise of an exploitable trade-off between inflation and unemployment; and that under Volcker's leadership, the Fed came to its senses, accepted the natural rate hypothesis, and focused monetary policy on setting inflation low.

Aspects of this "Berkeley view" receive backing from statistical work by Clarida, Gali, and Gertler (2000) and Taylor (1993), who fit monetary policy rules for subperiods that they choose to illuminate possible differences between the Burns and the

---

<sup>1</sup>By atheoretical we mean that the model's parameters are not explicitly linked to parameters describing decision makers' preferences and constraints.

Volcker-Greenspan eras. They find evidence for a systematic change of monetary policy across the two eras, a change that in Clarida, Gali, and Gertler’s ‘new-neoclassical-synthesis’ macroeconomic model would lead to better inflation-unemployment outcomes.

But Taylor’s and Clarida, Gertler, and Gali’s interpretation of the data has been disputed by Sims (1980, 1999) and Bernanke and Mihov (1998a, 1998b), both of whom have presented evidence that the U.S. data do not prompt rejection of the time invariance of the autoregressive coefficients of a VAR. They also present evidence for shifts in the variances of the innovations to their VARs. If one equation of the VAR is interpreted as describing a monetary policy rule, then Sims’s and Bernanke and Mihov’s results say that it was not the monetary policy strategy but luck (i.e., the volatility of the shocks) that changed between the Burns and the non-Burns periods.

### **1.3 Inflation persistence and inferences about the natural rate**

The persistence of inflation plays an important role in some widely used empirical strategies for testing the natural rate hypothesis and for estimating the natural unemployment rate. As we shall see, inflation persistence also plays an important role in lending relevance to instruments for estimating monetary policy rules. Therefore, we use our statistical model to portray the evolving persistence of inflation. We define a measure of persistence based on the normalized spectrum of inflation at zero frequency, then present how this measure of persistence increased during the 1960s and 70s, then fell during the 1980s and 1990s.

### **1.4 Drifting coefficients and the Lucas Critique**

Drifting coefficients have been an important piece of unfinished business within macroeconomic theory since Lucas played them up in the first half of his 1976 Critique, but then ignored them in the second half.<sup>2</sup> In Appendix A, we revisit how drifting coefficients bear on the theory of economic policy in the context of recent ideas about self-confirming equilibria. This appendix provides background for a view that helps to bolster the time-invariance view of the data taken by Sims and Bernanke and Mihov.

### **1.5 Method**

We take a Bayesian perspective and report time series of posterior densities for various economically interesting functions of hyperparameters and hidden states. We use a Markov Chain Monte Carlo algorithm to compute posterior densities.

---

<sup>2</sup>See Sargent (1999) for more about this interpretation of the two halves of Lucas’s 1976 paper.

## 1.6 Organization

The remainder of this paper is organized as follows. Section 2 describes the basic statistical model that we use to develop empirical evidence. We consign to appendix B a detailed description of the Markov Chain Monte Carlo algorithm that we use to approximate the posterior density of our model. Section 3 reports our results, and section 4 concludes. Appendix A pursues a theme opened in the Lucas Critique about how drifting coefficient models bear on alternative theories of economic policy.

## 2 A Bayesian Vector Autoregression with Drifting Parameters and Stochastic Volatility

The object of Cogley and Sargent (2001) was to develop empirical evidence about the evolving law of motion for inflation and to relate the evidence to stories about changes in monetary policy rules. To that end, we fit a Bayesian vector autoregression for inflation, unemployment, and a short term interest rate. We introduced drifting VAR parameters, so that the law of motion could evolve, but assumed the VAR innovation variance was constant. Thus, our measurement equation was

$$y_t = X_t' \theta_t + \varepsilon_t, \quad (1)$$

where  $y_t$  is a vector of endogenous variables,  $X_t$  includes a constant plus lags of  $y_t$ , and  $\theta_t$  is a vector of VAR parameters. The residuals,  $\varepsilon_t$ , were assumed to be conditionally normal with mean zero and constant covariance matrix  $R$ .

The VAR parameters were assumed to evolve as driftless random walks subject to reflecting barriers,

$$p(\theta_1, \dots, \theta_T | Q) \propto \prod_{s=0}^{T-1} I(\theta_{s+1}) f(\theta_{s+1} | \theta_s, Q). \quad (2)$$

The unrestricted transition density  $f(\theta_{t+1} | \theta_t, Q)$  represents a driftless random walk,

$$f(\theta_{t+1} | \theta_t, Q) \sim N(\theta_t, Q), \quad (3)$$

so that  $\theta_t$  evolves as

$$\theta_t = \theta_{t-1} + v_t, \quad (4)$$

apart from the reflecting barrier. The innovation  $v_t$  is normal with mean zero and variance  $Q$ , and we allowed for correlation between the state and measurement innovations,  $cov(v_t, \varepsilon_t) = C$ .

The reflecting barrier is encoded in the indicator function,  $I(\theta_t)$ , which takes a value of 0 when the roots of the associated VAR polynomial are inside the unit circle and is equal to 1 otherwise. This is a stability condition for the VAR, reflecting

an *a priori* belief about the implausibility of explosive representations for inflation, unemployment, and real interest. The stability prior follows from our belief that the Fed chooses policy rules in a purposeful way. Assuming that the Fed has a loss function that penalizes the variance of inflation, it will not choose a policy rule that results in a unit root in inflation, for that results in an infinite loss.<sup>3</sup>

In appendix C, we show that the joint density in equation 2 implies the transition density<sup>4</sup>

$$p(\theta_{t+1}|\theta_t, Q) \propto I(\theta_{t+1})f(\theta_{t+1}|\theta_t, Q)\pi(\theta_{t+1}, Q). \quad (5)$$

The term  $\pi(\theta_{t+1}, Q)$  represents the probability that random walk paths emanating from  $\theta_{t+1}$  will remain in the nonexplosive region going forward in time through the end of the sample. In other words, the transition density censors explosive draws from  $f(\theta_{t+1}|\theta_t, Q)$  and down-weights those likely to become explosive.<sup>5</sup>

## 2.1 Sims's and Stock's criticisms

Sims (2001) and Stock (2001) were concerned that our methods might exaggerate the time variation in  $\theta_t$ . One comment concerned the distinction between filtered and smoothed estimates. Cogley and Sargent (2001) reported results based on filtered estimates, and Sims pointed out that there is transient variation in filtered estimates even in time-invariant systems. In this paper, we report results based on smoothed estimates of  $\theta$ .

More importantly, Sims and Stock questioned our assumption that  $R$  is constant. They pointed to evidence developed by Bernanke and Mihov (1998a,b), Kim and Nelson (1999), McConnell and Perez Quiros (2000), and others that VAR innovation variances have changed over time. Bernanke and Mihov focused on monetary policy rules and found a dramatic increase in the variance of monetary policy shocks between

---

<sup>3</sup>To take a concrete example, consider the model of Rudebusch and Svensson (1999). Their model consists of an IS curve, a Phillips curve, and a monetary policy rule, and they endow the central bank with a loss function that penalizes inflation variance. The Phillips curve has adaptive expectations with the natural rate hypothesis being cast in terms of Solow and Tobin's unit-sum-of-the weights form. That form is consistent with rational expectations only when there is a unit root in inflation. The autoregressive roots for the system are not, however, determined by the Phillips curve alone; they also depend on the choice of monetary policy rule. With an arbitrary policy rule, the autoregressive roots can be inside, outside, or on the unit circle, but they are stable under optimal or near-optimal policies. When a shock moves inflation away from its target, poorly chosen policy rules may let it drift, but well-chosen rules pull it back.

<sup>4</sup>This formula supercedes the expression for the transition density given in Cogley and Sargent (2001). We are grateful to Simon Potter for pointing out an error in our earlier work and for suggesting ways to correct it.

<sup>5</sup>The probability that random walk trajectories will leave the nonexplosive region increases with the distance between  $t$  and  $T$ , but this tendency for  $\pi(\theta_{t+1}, Q)$  to decrease also affects the normalizing constant for equation (5). What matters is the relative likelihood of future instability, not the absolute likelihood.

1979 and 1982. Kim and Nelson and McConnell and Perez Quiros studied the growing stability of the U.S. economy, which they characterize in terms of a large decline in VAR innovation variances after the mid-1980s. The reason for this decline is the subject of debate, but there is now much evidence against our assumption of constant  $R$ .

Sims and Stock also noted that there is little evidence in the literature to support our assumption of drifting  $\theta$ . Bernanke and Mihov, for instance, used a procedure developed by Andrews (1993) to test for shifts in VAR parameters and were unable to reject time invariance. Indeed, their preferred specification was the opposite of ours, with constant  $\theta$  and varying  $R$ .

If the world were characterized by constant  $\theta$  and drifting  $R$ , and we fit an approximating model with constant  $R$  and drifting  $\theta$ , then it seems likely that our estimates of  $\theta$  would drift to compensate for misspecification of  $R$ , thus exaggerating the time variation in  $\theta$ . Stock suggested that this might account for our evidence on changes in inflation persistence. There is much evidence to support a positive relation between the level and variance of inflation, but the variance could be high either because of large innovation variances or because of strong shock persistence. A model with constant  $\theta$  and drifting  $R$  would attribute the high inflation variance of the 1970s to an increase in innovation variances, while a model with drifting  $\theta$  and constant  $R$  would attribute it to an increase in shock persistence. If Bernanke and Mihov are right, the evidence on inflation persistence reported in Cogley and Sargent (2001) paper may be an artifact of model misspecification.

## 2.2 Strategy for sorting out the issues

Of course, it is possible that both the coefficients and the volatilities vary, but most empirical models focus on one or the other. In this paper, we develop an empirical model that allows both to vary. We use the model to consider the extent to which drift in  $R$  undermines our evidence on drift in  $\theta$ , and also to conduct power simulations for the Andrews-Bernanke-Mihov test. Their null hypothesis, which they were unable to reject, was that  $\theta$  is time invariant. Whether this constitutes damning evidence against our vision of the world depends on the power of the test. Their evidence would be damning if the test reliably rejected a model like ours, but not so damning otherwise.

To put both elements in motion, we retain much of the specification described above, but now we assume that the VAR innovations can be expressed as

$$\varepsilon_t = R_t^{1/2} \xi_t, \tag{6}$$

where  $\xi_t$  is a standard normal random vector. Because we are complicating the model by introducing a drifting innovation variance, we simplify in another direction to economize on free parameters. Thus, we also assume that standardized VAR

innovations are independent of parameter innovations,

$$E(\xi_t v_s) = 0 \quad \text{for all } t, s. \quad (7)$$

To model drifting variances, we adopt a multivariate version of the stochastic volatility model of Jacquier, Polson, and Rossi (1994).<sup>6</sup> In particular, we assume that  $R_t$  can be expressed as

$$R_t = B^{-1} H_t B^{-\prime}, \quad (8)$$

where  $H_t$  is diagonal and  $B$  is lower triangular,

$$H_t = \begin{pmatrix} h_{1t} & 0 & 0 \\ 0 & h_{2t} & 0 \\ 0 & 0 & h_{3t} \end{pmatrix}, \quad (9)$$

$$B = \begin{pmatrix} 1 & 0 & 0 \\ \beta_{21} & 1 & 0 \\ \beta_{31} & \beta_{32} & 1 \end{pmatrix}. \quad (10)$$

The diagonal elements of  $H_t$  are assumed to be independent, univariate stochastic volatilities that evolve as driftless, geometric random walks,

$$\ln h_{it} = \ln h_{it-1} + \sigma_i \eta_{it}. \quad (11)$$

The random walk specification is designed for permanent shifts in the innovation variance, such as those emphasized in the literature on the growing stability of the U.S. economy. The volatility innovations,  $\eta_{it}$ , are standard normal random variables that are independent of one another and of the other shocks in the model,  $\xi_t$  and  $v_t$ . The volatility innovations are each scaled by a free parameter  $\sigma_i$  that determines their magnitude. The factorization in (8) and log specification in (11) guarantee that  $R_t$  is positive definite. The free parameters in  $B$  allow for correlation among the elements of  $\varepsilon_t$ . The matrix  $B$  orthogonalizes  $\varepsilon_t$ , but it is not an identification scheme.

This specification differs from others in the literature that assume finite-state Markov representations for  $R_t$ . Our specification has advantages and disadvantages relative to hidden Markov models. One advantage of the latter is that they permit jumps, whereas our model forces the variance to adjust continuously. An advantage of our specification is that it permits recurrent, permanent shifts in variance. Markov representations in which no state is absorbing permit recurrent shifts, but the system forever switches between the same configurations. Markov representations with an absorbing state permit permanent shifts in variance, but such a shift can only occur

---

<sup>6</sup>This formulation is closely related to the multi-factor stochastic volatility models of Aguilar and West (2001), Jacquier, Polson, and Rossi (1999), and Pitt and Shephard (1999).

once. Our specification allows permanent shifts to recur and allows new patterns to develop going forward in time.

We use Markov Chain Monte Carlo (MCMC) methods to simulate the posterior density.<sup>7</sup> Let

$$Y^T = [y'_1, \dots, y'_T]' \quad (12)$$

$$\theta^T = [\theta'_1, \dots, \theta'_T]' \quad (13)$$

and

$$H^T = \begin{bmatrix} h_{11} & h_{21} & h_{31} \\ h_{12} & h_{22} & h_{32} \\ \dots & \dots & \dots \\ h_{1T} & h_{2T} & h_{3T} \end{bmatrix} \quad (14)$$

represent the history of data, VAR parameters, and stochastic volatilities up to date  $T$ , let  $\sigma = (\sigma_1, \sigma_2, \sigma_3)$  stand for the standard deviations of the log-volatility innovations, and let  $\beta = [\beta_{21}, \beta_{31}, \beta_{32}]$  represent the free parameters in  $B$ . The posterior density,

$$p(\theta^T, Q, \sigma, \beta, H^T | Y^T), \quad (15)$$

summarizes beliefs about the model's free parameters, conditional on priors and the history of observations,  $Y^T$ .

## 3 Empirical Results

### 3.1 Data

In order to focus on the influence of drift in  $R$ , we use the same data as in our earlier paper. Inflation is measured by the CPI for all urban consumers, unemployment by the civilian unemployment rate, and the nominal interest rate by the yield on 3-month Treasury bills. Inflation and unemployment data are quarterly and seasonally adjusted, and Treasury bill data are the average of daily rates in the first month of each quarter. The sample spans the period 1948.1 to 2000.Q4. We work with VAR(2) representations for nominal interest, inflation, and the logit of unemployment.

### 3.2 Priors

The hyperparameters and initial states are assumed to be independent across blocks, so that the joint prior can be expressed as the product of marginal priors,

$$\begin{aligned} & p(\theta_0, h_{10}, h_{20}, h_{30}, Q, \beta, \sigma_1, \sigma_2, \sigma_3) \\ & = p(\theta_0)p(h_{10})p(h_{20})p(h_{30})p(Q)p(\beta)p(\sigma_1)p(\sigma_2)p(\sigma_3). \end{aligned} \quad (16)$$

---

<sup>7</sup>See appendix B for details.



Our prior for  $\theta_0$  is a truncated Gaussian density,

$$p(\theta_0) \propto I(\theta_0)f(\theta_0) = I(\theta_0)N(\bar{\theta}, \bar{P}). \quad (17)$$

The mean and variance of the Gaussian piece are calibrated by estimating a time-invariant vector autoregression using data for 1948.Q3-1958.Q4. The mean,  $\bar{\theta}$ , is set equal to the point estimate, and the variance,  $\bar{P}$ , is its asymptotic variance. Because the initial estimates are based on a short stretch of data, the location of  $\theta_0$  is only weakly restricted.

The matrix  $Q$  is a key parameter because it governs the rate of drift in  $\theta$ . We adopt an informative prior for  $Q$ , but we set its parameters to maximize the weight that the posterior puts on sample information. Our prior for  $Q$  is inverse-Wishart,

$$p(Q) = IW(\bar{Q}^{-1}, T_0), \quad (18)$$

with degrees of freedom  $T_0$  and scale matrix  $\bar{Q}$ . The degrees of freedom  $T_0$  must exceed the dimension of  $\theta_t$  in order for this to be proper. To put as little weight as possible on the prior, we set

$$T_0 = \dim(\theta_t) + 1. \quad (19)$$

To calibrate  $\bar{Q}$ , we assume

$$\bar{Q} = \gamma^2 \bar{P} \quad (20)$$

and set  $\gamma^2 = 3.5\text{e-}04$ . This makes  $\bar{Q}$  comparable to the value used in Cogley and Sargent (2001).<sup>8</sup> This setting can be interpreted as a weak version of a ‘business as usual’ prior, in the sense of Leeper and Zha (2001a,b). The prior is weak because it involves minimal degrees of freedom. It reflects a business-as-usual perspective because the implied values for  $\bar{Q}$  result in little variation in  $\theta$ . Indeed, had we calibrated  $Q = \bar{Q}$ , or set  $T_0$  so that a substantial weight was put on the prior, drift in posterior estimates of  $\theta$  would be negligible. Thus, the setting for  $\bar{Q}$  is conservative for our vision of the world.

The parameters governing priors for  $R_t$  are set more or less arbitrarily, but also very loosely, so that the data are free to speak about this feature as well. The prior for  $h_{i0}$  is log-normal,

$$p(\ln h_{i0}) = N(\ln \bar{h}_i, 10), \quad (21)$$

where  $\bar{h}_i$  is the initial estimate of the residual variance of variable  $i$ . Notice that a variance of 10 is huge on a natural log scale, making this weakly informative for  $h_{i0}$ . Similarly, the prior for  $\beta$  is normal with a large variance,

$$p(\beta) = N(0, 10000 \cdot I_3). \quad (22)$$

---

<sup>8</sup>An earlier draft experimented with alternative values of  $\gamma$  that push  $\bar{Q}$  toward zero, i.e. in the direction of less variation in  $\theta$ . We found only minor sensitivity to changes in  $\gamma$ .

Finally, the prior for  $\sigma_i^2$  is inverse gamma with a single degree of freedom,

$$p(\sigma_i^2) = IG\left(\frac{.01^2}{2}, \frac{1}{2}\right). \quad (23)$$

The specification is designed to put a heavy weight on sample information.

### 3.3 Details of the Simulation

We executed 100,000 replications of a Metropolis-within-Gibbs sampler and discarded the first 50,000 to allow for convergence to the ergodic distribution. We checked convergence by inspecting recursive mean plots of various parameters and by comparing results across parallel chains starting from different initial conditions. Because the output files are huge, we saved every 10th draw from the Markov chain, to economize on storage space. This has a side benefit of reducing autocorrelation across draws, but it does increase the variance of ensemble averages from the simulation. This yields a sample of 5000 draws from the posterior density. The estimates reported below are computed from averages of this sample.

### 3.4 The Posterior Mean of $Q$

We begin with evidence on the rate of drift in  $\theta$ , as summarized by posterior estimates of  $Q$ . Recall that  $Q$  is the virtual innovation variance for VAR parameters. Large values mean rapid movements in  $\theta$ , smaller values imply a slower rate of drift, and  $Q = 0$  represents a time-invariant model. The following table addresses two questions, whether the results are sensitive to the VAR ordering and how the stability prior influences the rate of drift in  $\theta$ .

**Table 1: Posterior Mean Estimates of  $Q$**

VAR Orderings	Stability Imposed		Stability Not Imposed	
	tr( $Q$ )	max( $\lambda$ )	tr( $Q$ )	max( $\lambda$ )
$i, \pi, u$	0.055	0.025	0.056	0.027
$i, u, \pi$	0.047	0.023	0.059	0.031
$\pi, i, u$	0.064	0.031	0.082	0.044
$\pi, u, i$	0.062	0.031	0.088	0.051
$u, i, \pi$	0.057	0.026	0.051	0.028
$u, \pi, i$	0.055	0.024	0.072	0.035

Note: The headings tr( $Q$ ) and max( $\lambda$ ) refer to the trace of  $Q$  and to the largest eigenvalue.

Sims (1980) reported that the ordering of variables in an identified VAR mattered for a comparison of interwar and postwar business cycles. In particular, for one ordering he found minimal changes in the shape of impulse response functions, with most of the difference between interwar and postwar cycles being due to a reduction in shock variances. He suggested to us that the ordering of variables might matter in our model too because of the way VAR innovation variances depend on the stochastic volatilities. In our specification, the first and second variables share common sources of stochastic volatility with the other variables, but the third variable has an independent source of volatility. Shuffling the variables might alter estimates of VAR innovation variances.

Accordingly, we estimated all possible orderings to see whether there exists an ordering that mutes evidence for drift in  $\theta$ , as in Sims (1980). This seems not to be the case. With the stability condition imposed (our preferred specification), there are only minor differences in posterior estimates of  $Q$ . The ordering that minimizes the rate of drift in  $\theta$  is  $[i_t, u_t, \pi_t]'$ , and the remainder of the paper focuses on this specification. This is conservative for our perspective, but results for the other orderings are similar.

The second question concerns how the stability prior influences drift in  $\theta$ . One might conjecture that the stability constraint amplifies evidence for drift in  $\theta$  by pushing the system away from the unit root boundary, forcing the model to fit inflation persistence via shifts in the mean. Again, this seems not to be the case; posterior mean estimates for  $Q$  are smaller when the stability condition is imposed. Withdrawing the stability prior increases the rate of drift in  $\theta$ .

The next table explores the structure of drift in  $\theta$ , focusing on the minimum- $Q$  ordering  $[i, u, \pi]'$ . Sargent's (1999) learning model predicts that reduced form parameters should drift in a highly structured way, because of the cross-equation restrictions associated with optimization and foresight. A formal treatment of cross-equation restrictions with parameter drift is a priority for future work. Here we report some preliminary evidence based on the principal components of  $Q$ .

**Table 2: Principal Components of  $Q$**

	Variance	Percent of Total Variation
1st PC	0.0230	0.485
2nd PC	0.0165	0.832
3rd PC	0.0054	0.945
4th PC	0.0008	0.963
5th PC	0.0007	0.978

Note: The second column reports the variance of the  $n$ th component (the  $n$ th eigenvalue of  $Q$ ), and the third states the fraction of the total variation (trace of  $Q$ ) for which the first  $n$  components account. The results refer to the minimum- $Q$  ordering  $[i, u, \pi]'$ .

The table confirms that drift in  $\theta$  is highly structured. There are 21 free parameters in a trivariate VAR(2) model, but only three linear combinations vary significantly over time. The first principal component accounts for almost half the total variation, the first two components jointly account for more than 80 percent, and the first three account for roughly 95 percent. These components load most heavily on lags of nominal interest and unemployment in the inflation equation; they differ in the relative weights placed on various lags. The remaining principal components, and the coefficients in the nominal interest and unemployment equations, are approximately time invariant. Thus the model's departure from time invariance is not as great as it first may seem. There are two or three drifting components in  $\theta$  that manifest themselves in a variety of ways.

### 3.5 The Evolution of $R_t$

Next we consider evidence on the evolution of  $R_t$ . Figure 1 depicts the posterior mean of  $R_t$  for the minimal- $Q$  ordering  $[i, u, \pi]'$ . The left-hand column portrays standard deviations for VAR innovations, expressed in basis points at quarterly rates, and the right-hand column shows correlation coefficients.

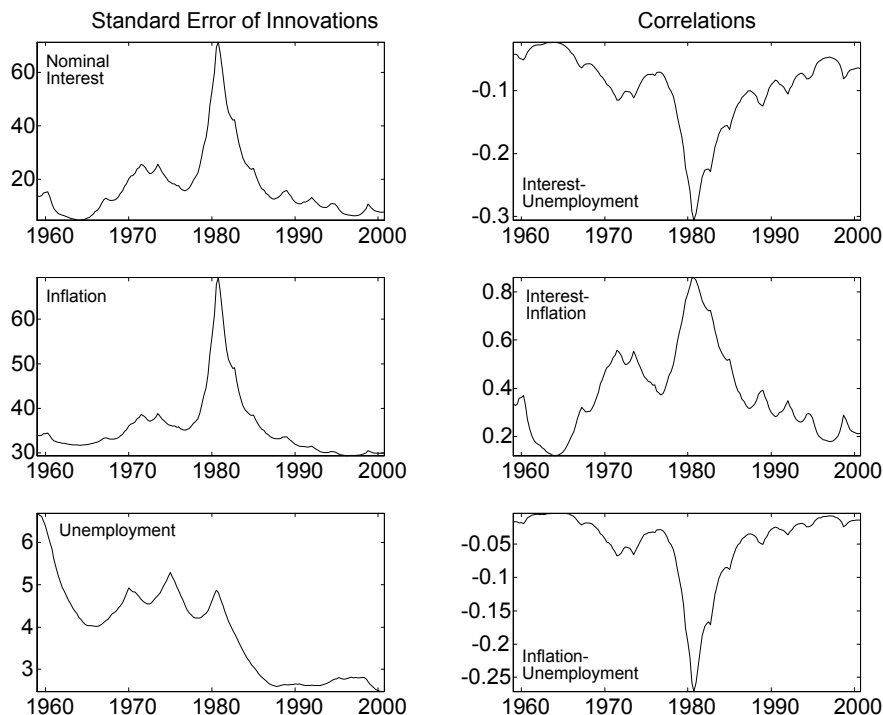


Figure 1: Posterior Mean of  $R_t$

The estimates support the contention that variation in  $R_t$  is an important feature of the data. Indeed, the patterns shown here resemble those reported by Bernanke and Mihov, Kim and Nelson, McConnell and Perez Quiros, and others.

For example, there is a substantial reduction in the innovation variance for unemployment in the early 1980s. At that time, the standard deviation fell by roughly 40 percent, an estimate comparable to those of Kim and Nelson and McConnell and Perez Quiros. Indeed, this seems to be part of a longer-term trend of growing stability in unemployment innovations. Our estimates suggest that there was a comparable decrease in variance in the early 1960s and that the standard error has fallen by a total of roughly 60 percent since the late 1950s. The trend toward greater stability was punctuated in the 1970s and early 1980s by countercyclical increases in variance. Whether the downward drift or business cycle pattern are likely to recur is an open question.

In addition, between 1979 and 1981, there is a spike in the innovation variances for nominal interest and inflation. The spike in the innovation variance for nominal interest resembles the estimates of Bernanke and Mihov. The two variances fell sharply after 1981 and reverted within a few years to levels achieved in the 1960s.

The right-hand column illustrates the evolution of correlations among the VAR innovations, calculated from the posterior mean,  $E(R_{t|T})$ . Unemployment innovations were negatively correlated with innovations in inflation and nominal interest throughout the sample. The correlations were largest in magnitude during the Volcker disinflation. At other times, the unemployment innovation was virtually orthogonal to the others. Inflation and nominal interest innovations were positively correlated throughout the sample, with the maximum degree of correlation again occurring in the early 1980s.

This correlation pattern has some bearing on one strategy for identifying monetary policy shocks. McCallum (1999) has argued that monetary policy rules should be specified in terms of lagged variables, on the grounds that the Fed lacks good current-quarter information about inflation, unemployment, and other target variables. This is especially relevant for decisions early in the quarter. If the Fed's policy rule depends only on lagged information, then it can be cast as the nominal interest equation in a VAR. Among other things, this means that nominal interest innovations are policy shocks and that correlations among VAR innovations represent unidirectional causation from policy shocks to the other variables.

The signs of the correlations in figure 1 suggest that this interpretation is problematic for our VAR. If nominal interest innovations were indeed policy shocks, conventional wisdom suggests they should be inversely correlated with inflation and positively correlated with unemployment, the opposite of what we find. A positive correlation with inflation and a negative correlation with unemployment suggests a policy reaction. There must be some missing information.<sup>9</sup>

Finally, figure 2 reports the total prediction variance,  $\log |E(R_{t|T})|$ . Following

---

<sup>9</sup>Two possibilities come to mind. There may be omitted lagged variables, so that the nominal interest innovation contains a component that is predictable based on a larger information set. The Fed may also condition on current-quarter reports of commodity prices or long term bond yields that are correlated with movements in inflation or unemployment.

Whittle (1953), we interpret this as a measure of the total uncertainty entering the system at each date.

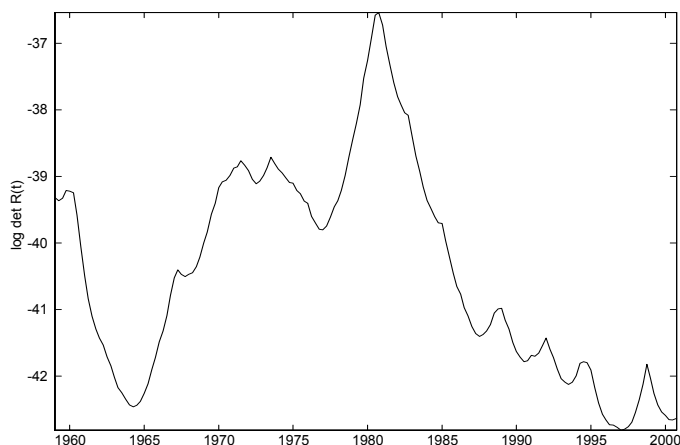


Figure 2: Total Prediction Variance

The smoothed estimates shown here are similar to the filtered estimates reported in our earlier paper. Both suggest a substantial increase in short-term uncertainty between 1965 and 1981 and an equally substantial decrease thereafter. The increase in uncertainty seems to have happened in two steps, one occurring between 1964 and 1972 and the other between 1977 and 1981. Most of the subsequent decrease occurred in the mid-1980s, during the latter years of Volcker’s term. This picture suggests that the growing stability of the economy may reflect a return to stability, though the earlier period of stability proved to be short-lived.

### 3.6 The Evolution of $\theta_t$

There is no question that variation in  $R$  is an interesting and important feature of the data, but does it alter the patterns of drift in  $\theta$  documented in our earlier paper? Our main interests concern movements in core inflation, the natural rate of unemployment, inflation persistence, the degree of policy activism, and how they relate to one another. Our interest in these features follows from their role in stories about how changes in monetary policy may have contributed to the rise and fall of inflation in the 1970s and 1980s.

#### 3.6.1 Core Inflation and the Natural Rate of Unemployment

The first set of figures depicts movements in core inflation and the natural rate of unemployment, which are estimated from local linear approximations to mean inflation

and unemployment, evaluated at the posterior mean,  $E(\theta_{t|T})$ . Write (1) in companion form as

$$z_t = \mu_{t|T} + A_{t|T}z_{t-1} + u_t, \quad (24)$$

where  $z_t$  consists of current and lagged values of  $y_t$ ,  $\mu_{t|T}$  contains the intercepts in  $E(\theta_{t|T})$ , and  $A_{t|T}$  contains the autoregressive parameters. By analogy with a time-invariant model, mean inflation at  $t$  can be approximated by

$$\bar{\pi}_t = s_\pi(I - A_{t|T})^{-1}\mu_{t|T}, \quad (25)$$

where  $s_\pi$  is a row vector that selects inflation from  $z_t$ . Similarly, mean unemployment can be approximated as

$$\bar{u}_t = s_u(I - A_{t|T})^{-1}\mu_{t|T}, \quad (26)$$

where  $s_u$  selects unemployment from  $z_t$ .

Figure 3 portrays the evolution of  $\bar{\pi}_t$  and  $\bar{u}_t$  for the ordering  $[i, u, \pi]'$ . Two features are worth noting. First, allowing for drift in  $R_t$  does not eliminate economically meaningful movements in core inflation or the natural rate. On the contrary, the estimates are similar to those in our earlier paper. Core inflation sweeps up from around 1.5 percent in the early 1960s, rises to a peak of approximately 8 percent in the late 1970s, and then falls to a range of 2.5 to 3.5 percent through most of the 1980s and 1990s. The natural rate of unemployment also rises in the late 1960s and 1970s and falls after 1980.

Second, it remains true that movements in  $\bar{\pi}_t$  and  $\bar{u}_t$  are highly correlated with one another, in accordance with the predictions of Parkin (1993) and Ireland (1999). The unconditional correlation is 0.748.

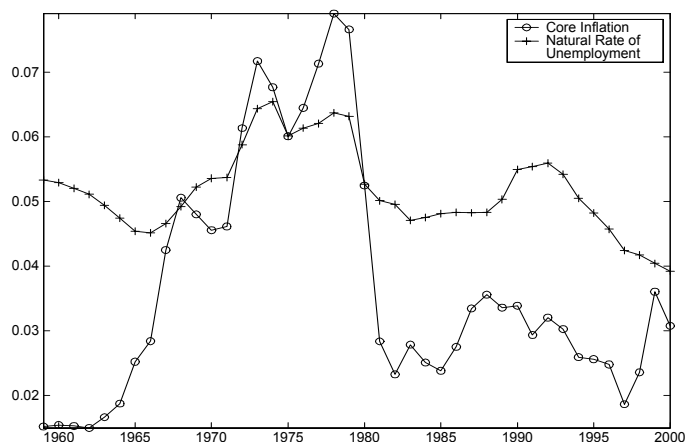


Figure 3: Core Inflation and the Natural Rate of Unemployment

Table 3 and figures 4 and 5 characterize the main sources of uncertainty about these estimates. The table and figures are based on a method developed by Sims and Zha (1999) for constructing error bands for impulse response functions. We start by estimating the posterior covariance matrix for  $\bar{\pi}_t$  via the delta method,

$$V_{\bar{\pi}} = \frac{\partial \bar{\pi}}{\partial \theta} V_{\theta} \frac{\partial \bar{\pi}}{\partial \theta}'. \quad (27)$$

$V_{\theta}$  is the  $KT \times KT$ <sup>10</sup> covariance matrix for  $\theta^T$  and  $\partial \bar{\pi} / \partial \theta$  is the  $T \times KT$  matrix of partial derivatives of the function that maps VAR parameters into core inflation, evaluated at the posterior mean of  $\theta^T$ . The posterior covariance  $V_{\theta}$  is estimated from the ensemble of Metropolis draws, and derivatives were calculated numerically.<sup>11</sup>

$V_{\bar{\pi}}$  is a large object, and we need a tractable way to represent the information it contains. Sims and Zha recommend error bands based on the first few principal components.<sup>12</sup> Let  $V_{\bar{\pi}} = W \Lambda W'$ , where  $\Lambda$  is a diagonal matrix of eigenvalues and  $W$  is an orthonormal matrix of eigenvectors. A two-sigma error band for the  $i$ th principal component is

$$\bar{\pi}_t \pm 2\lambda_i^{1/2} W_i, \quad (28)$$

where  $\lambda_i$  is the variance of the  $i$ th principal component and  $W_i$  is the  $i$ th column of  $W$ .

Table 3 reports the cumulative proportion of the total variation for which the principal components account. The second column refers to  $V_{\bar{\pi}}$ , and the third column decomposes the covariance matrix for the natural rate,  $V_{\bar{u}}$ . The other columns are discussed below.

One interesting feature is the number of non-trivial components. The first principal component in  $V_{\bar{\pi}}$  and  $V_{\bar{u}}$  accounts for 40 to 50 percent of the total variation, and the first 5 jointly account for about 75 percent. This suggests an important departure from time invariance. In a time-invariant model, there would be a single factor representing uncertainty about the location of the terminal estimate, but smoothed estimates going backward in time would be perfectly correlated with the terminal estimate and would contribute no additional uncertainty.<sup>13</sup>  $V$  would be a  $T \times T$  matrix

<sup>10</sup> $K$  is the number of elements in  $\theta$ , and  $T$  represents the number of years. We focused on every fourth observation to keep  $V$  to a manageable size.

<sup>11</sup>This roundabout method for approximating  $V_{\bar{\pi}}$  was used because the direct estimate was contaminated by a few outliers, which dominated the principal components decomposition on which Sims-Zha bands are based. The outliers may reflect shortcomings of our linear approximations near the unit root boundary.

<sup>12</sup>If the elements of  $\bar{\pi}_t$  were uncorrelated across  $t$ , it would be natural to focus instead on the diagonal elements of  $V_{\bar{\pi}}$ , e.g. by graphing the posterior mean plus or minus two standard errors at each date. But  $\bar{\pi}_t$  is serially correlated, and Sims and Zha argue that a collection of principal components bands better represents the shape of the posterior in such cases.

<sup>13</sup>Setting  $Q = 0$  in the Kalman filter implies  $P_{t+1|t} = P_{t|t}$ . Then the covariance matrix in the backward recursion of the Gibbs sampler would be  $P_{t|t+1} = 0$ , implying a perfect correlation between draws of  $\theta_{t+1}$  and  $\theta_t$ .



**Table 3: Principal Component Decomposition for Sims-Zha Bands**

	$V_{\bar{\pi}}$	$V_{\bar{u}}$	$V_{g_{\pi\pi}}$	$V_A$
1st PC	0.521	0.382	0.374	0.662
2nd PC	0.604	0.492	0.490	0.801
3rd PC	0.674	0.597	0.561	0.870
4th PC	0.715	0.685	0.612	0.906
5th PC	0.750	0.727	0.662	0.936
6th PC	0.778	0.767	0.701	0.949
8th PC	0.822	0.820	0.756	0.972
10th PC	0.851	0.856	0.800	0.984

Note: Entries represent the cumulative percentage of the total variation (trace of  $V$ ) for which the first  $n$  principal components account.

with rank one, and the single principal component would describe uncertainty about the terminal location. In a nearly time-invariant model, i.e. one with small  $Q$ , the path to the terminal estimate might wiggle a little, but one would still expect uncertainty about the terminal estimate to dominate. That the first component accounts for a relatively small fraction of the total suggests there is also substantial variation in the shape of the path.

Error bands for core inflation are shown in figure 4. The central dotted line is the posterior mean estimate, reproduced from figure 3. The horizontal line is a benchmark, end-of-sample, time-invariant estimate of mean inflation.

The first principal component, which accounts for roughly half the total variation, describes uncertainty about the location of core inflation in the late 1960s and 1970s. As core inflation increased, so too did uncertainty about the mean, and by the end of the decade a two-sigma band ranged from 2 to 14 percent. The growing uncertainty about core inflation seems to be related to changes in inflation persistence. Core inflation can be interpreted as a long-horizon forecast, and the variance of long-horizon forecasts depends positively on the degree of persistence. As shown below, inflation also became more persistent as core inflation rose. Indeed, our estimates of inflation persistence are highly correlated with the width of the first error band.

Components 3 through 5 portray uncertainty about the number of local peaks in the 1970s, and they jointly account for about 15 percent of the total variation. Bands for these components cross several times, a sign that some paths had more peaks than others. For example, in panel 3, trajectories associated with a global peak at the end of the 1970s tended also to have a local peak at the end of the 1960s. In contrast, paths that reached a global peak in the mid-1970s tended to have a single peak.

Finally, the sixth component loads heavily on the last few years in the sample, describing uncertainty about core inflation in the late 1990s. At the end of 2000, a two-sigma band for this component ranged from approximately 1 to 5 percent.

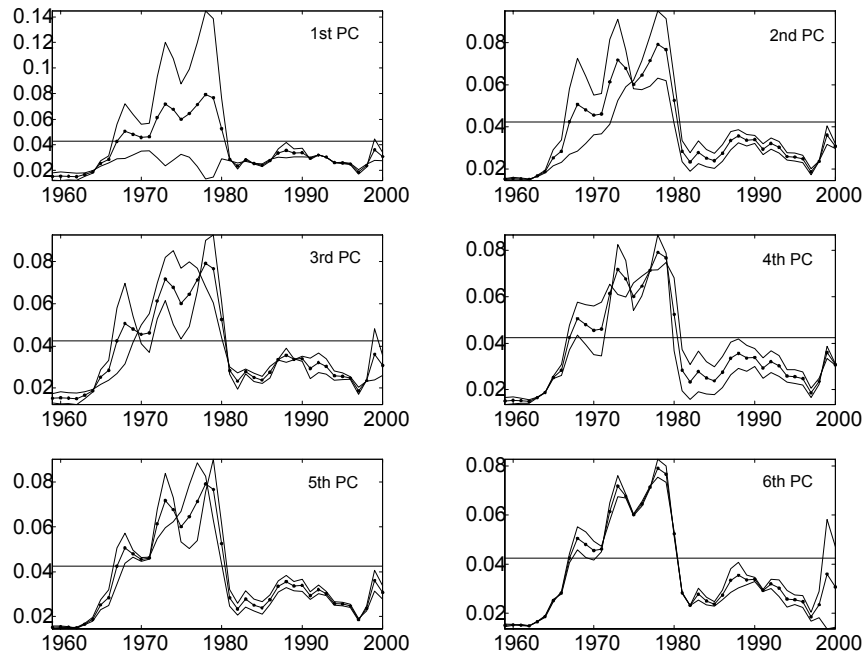


Figure 4: Two-Sigma Error Bands for Core Inflation

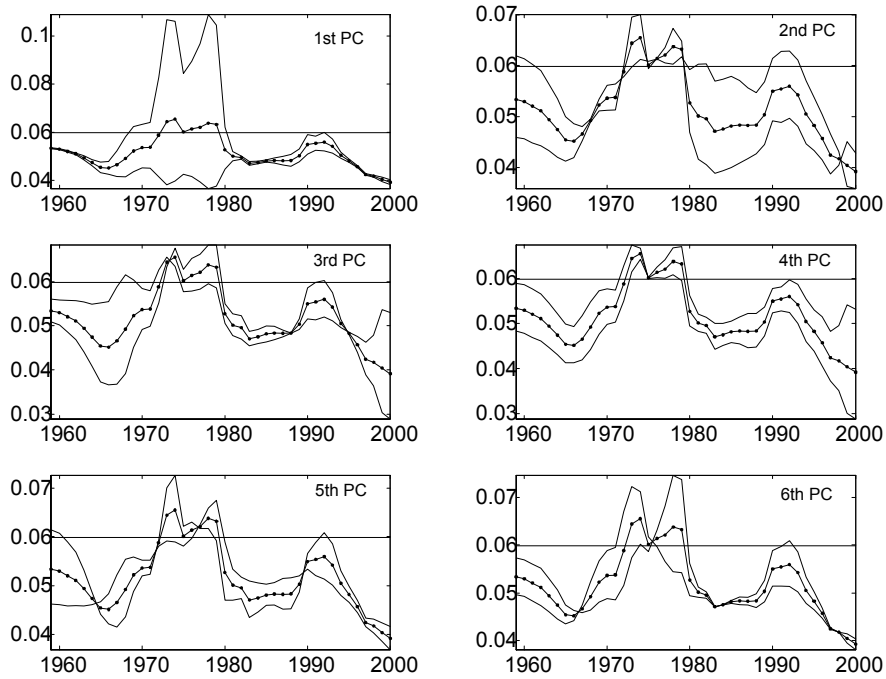


Figure 5: Two-Sigma Error Bands for the Natural Rate of Unemployment

Error bands for the natural rate are constructed in the same way, and they are shown in figure 5. Once again, the central dotted line is the posterior mean estimate, and the horizontal line is an end-of-sample, time-invariant estimate of mean unemployment. The first principal component in  $V_{\bar{u}}$  also characterizes uncertainty about the 1970s. The error band widens in the late 1960s when the natural rate began to rise, and it narrows around 1980 when the mean estimate fell. The band achieved its maximum width around the time of the oil shocks, when it ranged from roughly 4 to 11 percent. The width of this band also seems to be related to changes in the persistence of shocks to unemployment.

The second, third, and fourth components load heavily on the other years of the sample, jointly accounting for about 30 percent of the total variation. Roughly speaking, they cover intervals of plus or minus 1 percentage point around the mean. The fifth and sixth components account for 8 percent of the variation, and they seem to be related to uncertainty about the timing and number of peaks in the natural rate.

### 3.6.2 Inflation Persistence

Next we turn to evidence on the evolution of second moments of inflation. Second moments are measured by a local linear approximation to the spectrum for inflation,

$$f_{\pi\pi}(\omega, t) = s_{\pi}(I - A_{t|T}e^{-i\omega})^{-1} \frac{E(R_{t|T})}{2\pi} (I - A_{t|T}e^{i\omega})^{-1'} s_{\pi}'. \quad (29)$$

evaluated at the posterior mean of  $\theta$  and  $R$ . An estimate of  $f_{\pi\pi}(\omega, t)$  is shown in figure 6. Time is plotted on the  $x$ -axis, frequency on the  $y$ -axis, and power on the  $z$ -axis.

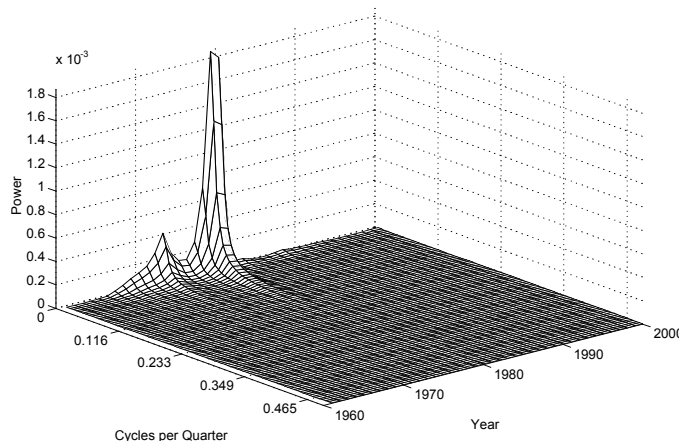


Figure 6: Spectrum for Inflation

Again, the estimates are similar to those reported in Cogley and Sargent (2001). The introduction of drift in  $R_t$  does not undermine our evidence on variation in the spectrum for inflation.

The most significant feature of this graph is the variation over time in the magnitude of low frequency power. In our earlier paper, we interpreted the spectrum at zero as a measure of inflation persistence. Here that interpretation is no longer quite right, because variation in low-frequency power depends not only on drift in the autoregressive parameters,  $A_{t|T}$ , but also on movements in the innovation variance,  $E(R_{t|T})$ . In this case, the normalized spectrum,

$$g_{\pi\pi}(\omega, t) = \frac{f_{\pi\pi}(\omega, t)}{\int_{-\pi}^{\pi} f_{\pi\pi}(\omega, t) d\omega}, \quad (30)$$

provides a better measure of persistence. The normalized spectrum is the spectrum divided by the variance in each year. The normalization adjusts for changes in innovation variances and measures autocorrelation rather than autocovariance. We interpret  $g_{\pi\pi}(0, t)$  as a measure of inflation persistence.

Estimates of the normalized spectrum are shown in figure 7. As in figure 6, the dominant feature is the variation over time in low-frequency power, though the variation in  $g_{\pi\pi}(0, t)$  differs somewhat from that in  $f_{\pi\pi}(0, t)$ . Instead of sharp spikes in the 1970s,  $g_{\pi\pi}(0, t)$  sweeps gradually upward in the latter half of the 1960s and remains high throughout the 1970s. The spectrum at zero falls sharply after 1980, and there is discernible variation throughout the remainder of the sample.

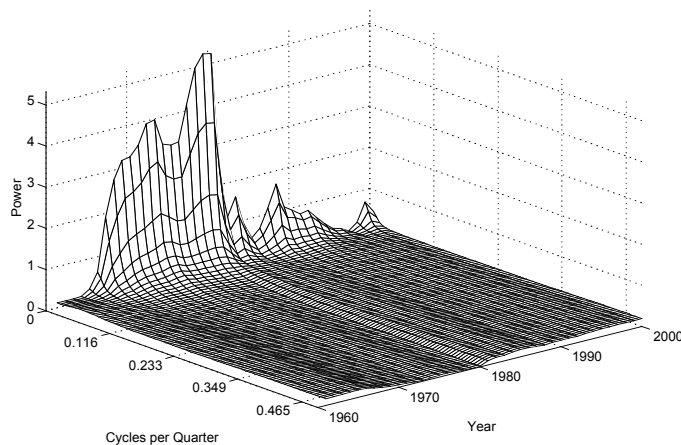


Figure 7: Normalized Spectrum for Inflation

Figure 8 depicts two-sigma error bands for  $g_{\pi\pi}(0, t)$ , based on the principal components of its posterior covariance matrix,  $V_{g_{\pi\pi}}$ . The latter was estimated in the same

way as  $V_{\bar{\pi}}$  or  $V_{\bar{u}}$ . The third column in table 3 indicates that the first component in  $V_{g_{\pi\pi}}$  accounts for only 37 percent of the total variation and that the first 5 components jointly account for 84 percent. Again, this signifies substantial variation in the shape of the path for  $g_{\pi\pi}(0, t)$ .

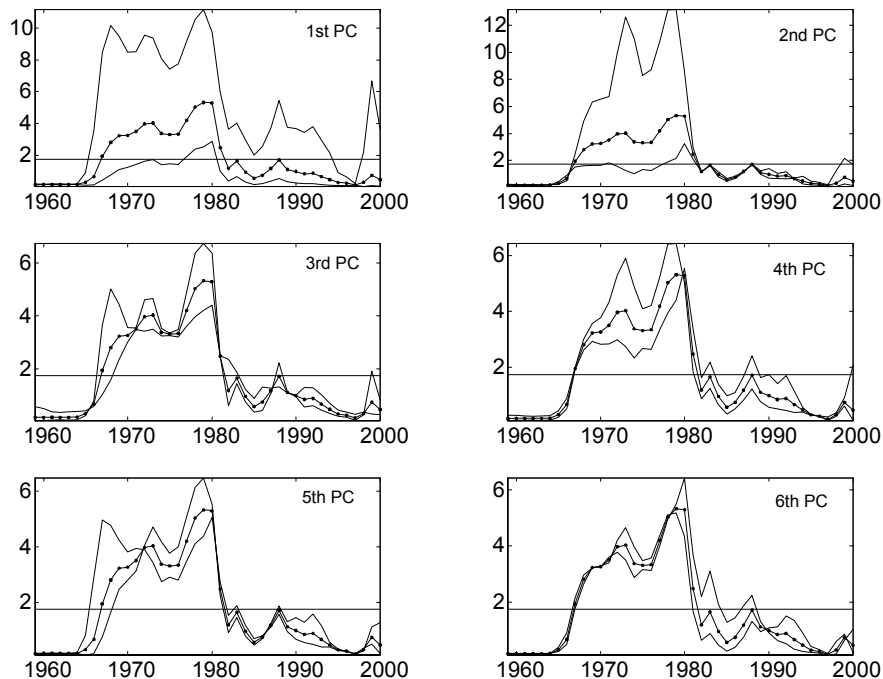


Figure 8: Two-Sigma Error Bands for the Normalized Spectrum At Zero

Error bands for the first two components load heavily on the 1970s. Although the bands suggest there was greater persistence than in the early 1960s or mid-1990s, the precise magnitude of the increase is hard to pin down. Roughly speaking, error bands for the first two components suggest that  $g_{\pi\pi}(0, t)$  was somewhere between 2 and 10. For the sake of comparison, a univariate AR(1) process with coefficients of 0.85 to 0.97 has values of  $g_{\pi\pi}(0)$  in this range. In contrast, the figure suggests that inflation was approximately white noise in the early 1960s and not far from white noise in the mid-1990s. Uncertainty about inflation persistence was increasing again at the end of the sample.

The third, fourth, and fifth components reflect uncertainty about the timing and number of peaks in  $g_{\pi\pi}(0, t)$ . For example, panels 3 and 5 suggest that paths on which there was a more gradual increase in persistence tended to have a big global peak in the late 1970s, while those on which there was a more rapid increase tended to have comparable twin peaks, first in the late 1960s and then again in 1980. Panel 4 suggests that some paths had twin peaks at the time of the oil shocks, while others had a single peak in 1980. These components jointly account for about 17 percent of the total variation.

One of the questions in which we are most interested concerns the relation between inflation persistence and core inflation. In Cogley and Sargent (2001), we reported evidence of a strong positive correlation. Here we also find a strong positive correlation, equal to 0.92.

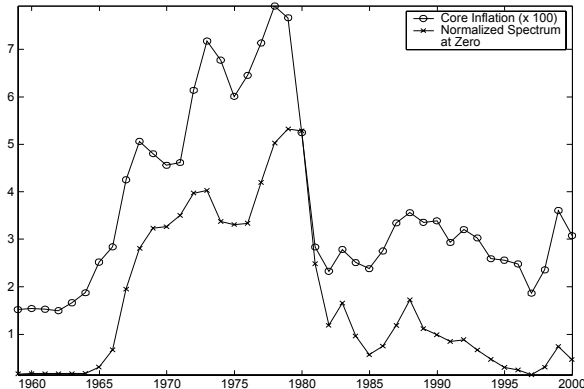


Figure 9: Core Inflation and Inflation Persistence

The relation between the two series is illustrated in figure 9, which reproduces estimates from figures 3 and 7. As core inflation rose in the 1960s and 1970s, inflation also became more persistent. Both features fell sharply during the Volcker disinflation. This correlation is problematic for the escape route models of Sargent (1999) and Cho, Williams, and Sargent (2002), which predict that inflation persistence grows along the transition from high to low inflation. Our estimates suggest the opposite pattern.

### 3.6.3 Monetary Policy Activism

Finally, we consider evidence on the evolution of policy activism. Following Clarida, Gali, and Gertler (2000), we estimate this from a forward-looking Taylor rule with interest smoothing,

$$i_t = \beta_0 + \beta_1 E_t \pi_{t,t+h_\pi} + \beta_2 E_t u_{t,t+h_u} + \beta_3 i_{t-1} + \eta_t, \quad (31)$$

where  $\pi_{t,t+h_\pi}$  represents average inflation from  $t$  to  $t + h_\pi$  and  $u_{t,t+h_u}$  is average unemployment. The activism parameter is defined as

$$A = \beta_1(1 - \beta_3)^{-1}, \quad (32)$$

and the policy rule is said to be activist if  $A \geq 1$ . With a Ricardian fiscal policy, an activist monetary rule delivers a determinate equilibrium. Otherwise, sunspots may matter for inflation and unemployment.

We interpret the parameters of the policy rule as projection coefficients and compute projections from our VAR. This is done via two-stage least squares on a date-by-date basis. The first step involves projecting the Fed's forecasts  $E_t \pi_{t,t+h_\pi}$  and  $E_t u_{t,t+h_u}$

onto a set of instruments, and the second involves projecting current interest rates onto the fitted values. At each date, we parameterize the VAR with posterior mean estimates of  $\theta_t$  and  $R_t$  and calculate population projections associated with those values.

The instruments chosen for the first-stage projection must be elements of the Fed's information set. Notice that a complete specification of their information set is unnecessary; a subset of their conditioning variables is sufficient for forming first-stage projections, subject of course to the order condition for identification. Among other variables, the Fed observes lags of inflation, unemployment, and nominal interest when making current-quarter decisions, and we project future inflation and unemployment onto a constant and two lags of each. Thus, our instruments for the Fed's forecasts  $E_t\pi_{t,t+h_\pi}$  and  $E_tu_{t,t+h_u}$  are the VAR forecasts  $E_{t-1}\pi_{t,t+h_\pi}$  and  $E_{t-1}u_{t,t+h_u}$ , respectively.<sup>14</sup>

Because  $\theta_t$  and  $R_t$  evolve over time, so too do the policy projections. Point estimates for  $A$  are shown in figure 10. Here we assume  $h_\pi = 4$  and  $h_u = 2$ , but the results for one-quarter ahead forecasts are similar.

The estimates broadly resemble those reported by Clarida, et. al., as well as those in our earlier paper. The estimated policy rule was activist in the early 1960s, but became approximately neutral in the late 1960s and then turned passive in the early 1970s. It remained passive until the early 1980s. The estimate of  $A$  rose sharply around the time of the Volcker disinflation, and it has remained in the activist region ever since. As shown in figure 11, the estimates of  $A$  are inversely related to core inflation and the degree of inflation persistence, suggesting that changes in the degree of activism may have contributed to the rise and fall of inflation.

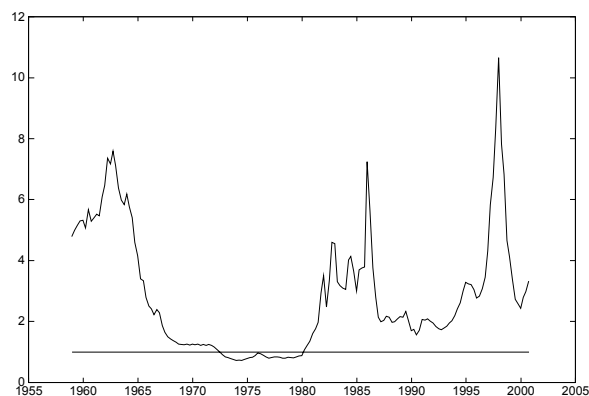


Figure 10: Estimates of the Activism Coefficient

---

<sup>14</sup>Here we follow McCallum, who warns against the assumption that the Fed sees current quarter inflation and unemployment when making decisions. This strategy also sidesteps assumptions about how to orthogonalize current quarter innovations. This is an important advantage of the Clarida, et. al. approach relative to structural VAR methods. Establishing that the Fed can observe *some* variables is easier than compiling a complete list of what the Fed sees.

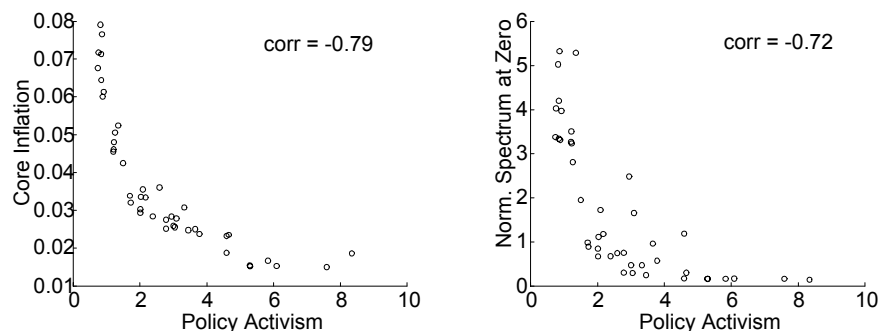


Figure 11: Policy Activism, Core Inflation, and Inflation Persistence

Figure 12 suggests, however, that some qualifications are necessary, especially at the beginning and end of the sample. The figure portrays two-sigma error bands based on the principal components of the posterior covariance matrix,  $V_A$ . The last column of table 3 shows that several principal components contribute to  $V_A$ , with the first component accounting for about two-thirds of the total variation. That there is more than one important component is evidence for variation in the path of  $A$ .

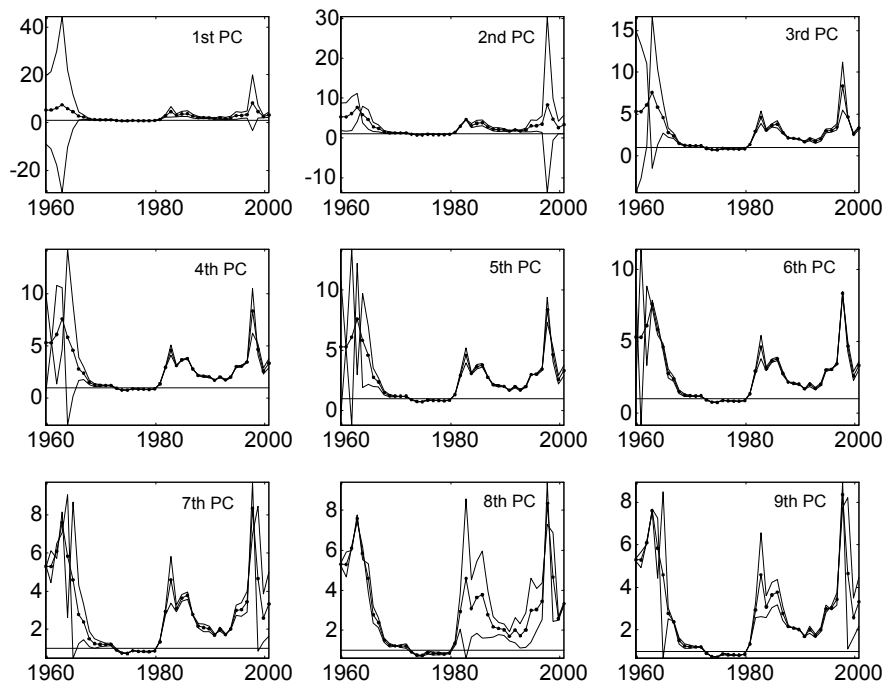


Figure 12: Two-Sigma Error Bands for the Activism Parameter

But the shape of the path is well determined only in the middle of the sample. The first four principal components record substantial uncertainty at the beginning and



end. We interpret this as a symptom of weak identification. Substantial uncertainty about  $A$  occurs at times when inflation is weakly persistent. Our instruments have little relevance when future inflation is weakly correlated with lagged variables, and the policy rule parameters are weakly identified at such times. Thus, inferences about  $A$  are fragile at the beginning and end of the sample. There is better evidence of changes in  $A$  during the middle of the sample. Lagged variables are more relevant as instruments for the 1970s, when inflation and unemployment were very persistent, and for that period the estimates are more precise.

The next figure characterizes more precisely how the posterior for  $A_t$  differs across the Burns and Volcker-Greenspan terms. It illustrates histograms for  $A_t$  for the years 1975, 1985, and 1995. The histograms were constructed by calculating an activism parameter for each draw of  $\theta_t$  and  $R_t$  in our simulation, for a total of 5000 in each year.<sup>15</sup> Values for 1975 are shown in black, those for 1985 are in white, and estimates for 1995 are shown in yellow.

In 1975, the probability mass was concentrated near 1, and the probability that  $A_t > 1$  was 0.208. By 1985, the center of the distribution had shifted to the right, and the probability that  $A_t > 1$  had increased to 0.919. The distribution for 1995 is

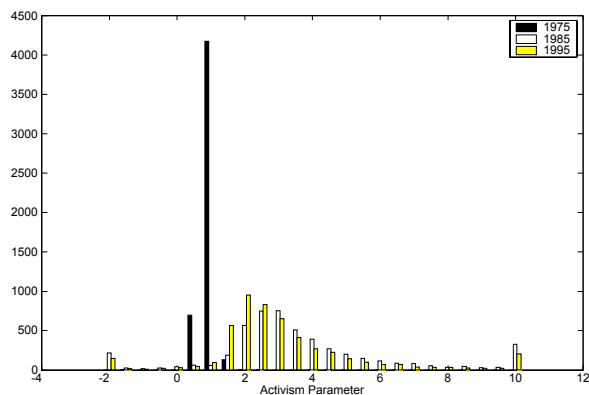


Figure 13: Histograms for  $A_t$  in Selected Years

similar to that for 1985, with a 0.941 probability that  $A_t > 1$ . Comparing estimates along the same sample paths, the probability that  $A_t$  increased between 1975 and 1985 is 0.923, and the probability that it increased between 1975 and 1995 is 0.943.

The estimates seem to corroborate those reported by Clarida, et. al. that monetary policy was passive in the 1970s and activist for much of the Volcker-Greenspan era. Estimates for the latter period are less precise, but it seems clear that the probability distribution for  $A_t$  shifted to the right.

<sup>15</sup>Outliers are collected in the end bins.

### 3.7 Tests for $\theta$ Stability

Finally, we consider classical tests for variation in  $\theta$ . Bernanke and Mihov (1998a,b) were also concerned about the potential for shifts in VAR parameters arising from changes in monetary policy, and they applied a test developed by Andrews (1993) to examine stability of  $\theta$ . For reduced form vector autoregressions similar to ours, they were unable to reject the hypothesis of time invariance.

We applied the same test to our data and found the same results. We considered two versions of Andrews’s *sup-LM* test, one that examines parameter stability for the VAR as a whole and another that tests stability on an equation-by-equation basis. The results are summarized in table 4. Columns labelled with variable names refer to single-equation tests, and the column labelled ‘VAR’ refers to a test for the system as a whole. In each case, we fail to reject that  $\theta$  is time invariant.<sup>16</sup>

Bernanke and Mihov correctly concluded that the test provides little evidence against stability of  $\theta$ . But does the result constitute evidence against parameter instability? A failure to reject provides evidence against an alternative hypothesis only if it has reasonably high power. Whether this test has high power against a model like ours is an open question, so we decided to investigate it.

**Table 4: Andrews’s *sup-LM* Test**

	Nominal Interest	Unemployment	Inflation	VAR
Data	F	F	F	F
Power	0.136	0.172	0.112	0.252

Note: An ‘F’ means the test fails to reject at the 10 percent level when applied to actual data. Entries in the second row refer to the fraction of artificial samples in which the null hypothesis is rejected at the 5 percent level.

To check the power of the test, we performed a Monte Carlo simulation using our drifting parameter VAR as a data generating process. To generate artificial data, we parameterized equation (1) with draws of  $\theta^T$ ,  $H^T$ , and  $B$  from the posterior density. For each draw of  $(\theta^T, H^T, B)$ , we generated an artificial sample for inflation, unemployment, and nominal interest and then calculated the *sup-LM* statistics. We performed 10,000 replications and counted the fraction of samples in which the null hypothesis of constant  $\theta$  is rejected at the 5 percent level. The results are summarized in the second row of table 4.

The power of the test is never very high. The VAR test has highest the success rate, detecting drift in  $\theta$  in about one-fourth of the samples. The detection probabilities are lower in the single equation tests, which reject at the 5 percent level in only

---

<sup>16</sup>We also performed a Monte Carlo simulation to check the size of the Andrews test; the results confirmed that size distortions do not explain the failure to reject.

about 14 percent of the samples. Thus, even when  $\theta$  drifts in the way we describe, a failure to reject is at least 3 times as likely as a rejection.

Andrews’s test is designed to have power against alternatives involving a single shift in  $\theta$  at some unknown break date. The results of this experiment may just reflect that this test is less well suited to detect alternatives such as ours that involve continual shifts in parameters. Accordingly, we also investigate a test developed by Nyblom (1989) and Hansen (1992) that is designed to have power against alternatives in which parameters evolve as driftless random walks. Results for the Nyblom-Hansen test are summarized in table 5.

**Table 5: The Nyblom-Hansen Test**

	Nominal Interest	Unemployment	Inflation	VAR
Data	F	F	F	F
Power	0.076	0.170	0.086	0.234

See the note to table 4.

When applied to actual data, the Nyblom-Hansen test also fails to reject time invariance for  $\theta$ . To examine its power, we conducted another Monte Carlo simulation using our drifting parameter VAR as a data generating mechanism, and we found that this test also has low power against our representation. Indeed, the detection probabilities are a bit lower than those for the *sup-LM* test.

Boivin (1999) conjectures that the *sup-Wald* version of Andrews’s test may have higher power than the others, and so we also consider this procedure. The results, which are shown in table 6, provide some support for his conjecture. The detection probability is higher in each case, and it is substantially higher for the inflation equation. Indeed, this is the only case among the ones we study in which the detection probability exceeds 50 percent. It is noteworthy that in this case we also strongly reject time invariance in the actual data. Time invariance is also rejected for the VAR as a whole.

**Table 6: Andrews’s sup-Wald Test**

	Nominal Interest	Unemployment	Inflation	VAR
Data	F	F	R 1%	R 5%
Power	0.173	0.269	0.711	0.296

Note: ‘R x%’ signifies a rejection at the x percent level.

We made two other attempts to concentrate power in promising directions. The first focuses on parameters of the Clarida, et. al. policy rule. If drift in  $\theta$  is indeed a manifestation of changes in monetary policy, then tests for stability of the latter

should be more powerful than for stability of the former. The vector  $\theta$  has high dimension, and the drifting components in  $\theta$  should lie in a lower-dimensional subspace corresponding to drifting policy parameters.<sup>17</sup> To test stability of the Clarida, et. al. rule, we estimated a version for the period 1959-2000 using our data and instruments, and calculated Andrews's statistics.<sup>18</sup> Perhaps surprisingly in light of their results, the tests fail to reject time invariance (see table 7). We repeated the procedure for artificial samples generated from our VAR to check the power of the test. Once again, the results show that the tests have low detection probabilities.

**Table 7: Stability of the CGG Policy Rule**

	sup-LM	sup-Wald
Data	F	F
Power	0.143	0.248

See the note to table 4.

We also tried to increase power by concentrating on a single linear combination of  $\theta$  that we think is most likely to vary. The linear combination with greatest variance is the first principal component, and we used the dominant eigenvector of the sample variance of  $E(\Delta\theta_{t|T})$  to measure this component. As figures 14 and 15 illustrate, the first principal component dominates the variation in  $\theta_{t|T}$ ;<sup>19</sup> most of the other principal components are approximately time invariant. The first component is also highly correlated with variation in the features discussed above. Thus it seems to be a promising candidate on which to concentrate.

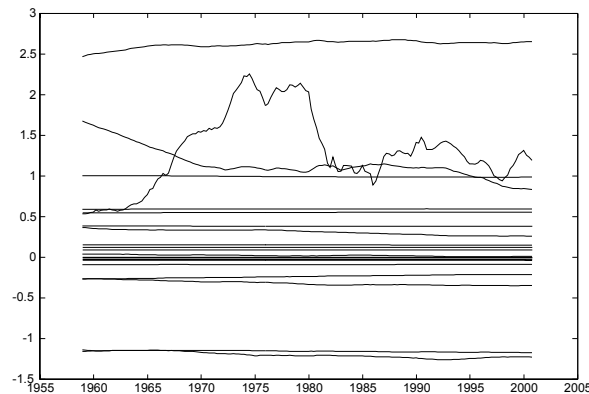


Figure 14: Principal Components of  $\theta_{t|T}$

<sup>17</sup>This assumes that shifts in policy are the only source of drift in  $\theta$ .

<sup>18</sup>We chose Andrews's tests because the CCG rule is estimated by GMM. The Nyblom-Hansen test is based on ML estimates.

<sup>19</sup>More precisely, the figures illustrate partial sums of the first principal component for  $\Delta\theta_{t|T}$ .

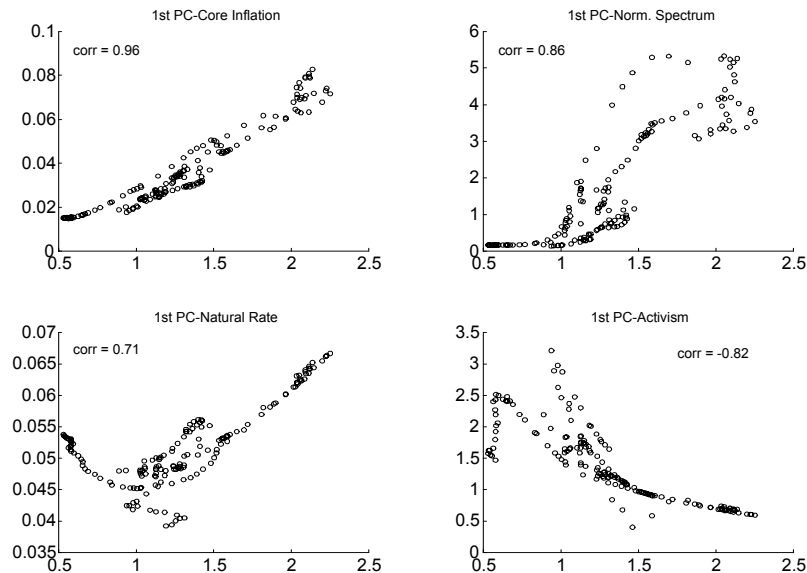


Figure 15: Correlation of First Principal Component with Other Features

Yet the results of a Monte Carlo simulation, shown in table 8, suggest that power remains low, with a rejection probability of only about 15 percent. Indeed, the procedure is inferior to the VAR tests reported above. Agnosticism about drifting components in  $\theta$  seems to be better. Despite the low power, one of the tests rejects time invariance in actual data.

**Table 8: Stability of the First Principal Component**

	<i>sup-LM</i>	<i>sup-Wald</i>
Data	F	R 5%
Power	0.220	0.087

See the note to table 4.

To summarize, most of our tests fail to reject time invariance of  $\theta$ , but most also have low power to detect the patterns of drift we describe above. In the one case where a test has a better-than-even chance of detecting drift in  $\theta$ , for the data time invariance is rejected at better than the one-percent level. One reasonable interpretation is that  $\theta$  is drifting, but that most of the procedures are unable to detect it.

Perhaps low power should not be a surprise. Our model nests the null of time invariance as a limiting case, i.e. when  $Q = 0$ . One can imagine indexing a family

of alternative models in terms of  $Q$ . For  $Q$  close to zero, size and power should be approximately the same. Power should increase as  $Q$  gets larger, and eventually the tests are likely to reject with high probability. But in between there is a range of alternative models, arrayed in terms of increasing  $Q$ , that the tests are unlikely to reject. The message of the Monte Carlo detection statistics is that a model such as ours with economically meaningful drift in  $\theta$  often falls in the indeterminate range.

## 4 Conclusion

One respectable view is that either an erroneous model, insufficient patience, or his inability to commit to a better policy made Arthur Burns respond to the end of Bretton Woods by administering monetary policy in a way that produced the greatest peace time inflation in U.S. history; and that an improved model, more patience, or greater discipline led Paul Volcker to administer monetary policy in a way that conquered American inflation.<sup>20</sup> Another respectable view is that what distinguished Burns and Volcker was not their models or policies but their luck. This paper and its predecessor (Cogley and Sargent (2001)) fit time series models that might help distinguish these views.

This paper also responds to Sims's (2001) and Stock's (2001) criticism of the evidence for drifting systematic parts of vector autoregressions in Cogley and Sargent (2001) by altering our specification to include stochastic volatility. While we have found evidence for drifting variances within our new specification, we continue to find evidence that the VAR coefficients have drifted, mainly along one important direction. Our model is atheoretical, but for reasons discussed in Appendix A and also by Sargent (1999) and Luca Benati (2001), the presence of drifting coefficients contains clues about whether government policy makers' models or preferences have evolved over time.

It is appropriate to be cautious in accepting evidence either for or against drifting coefficients. For reasons that are most clear in continuous time (see Anderson, Hansen, and Sargent (2000)), it is much more difficult to detect evidence for movements in the systematic part of a vector autoregression than it is to detect stochastic volatility. This situation is reflected in the results of our experiments with implementing Bernanke and Mihov's tests under an artificial economy with drifting coefficients.

## A Theories of economic policy

Contrasting visions of aggregate economic time series that we attribute to Lucas (1976), Sargent and Wallace (1976), and Sims (1982) can be represented within the following modification of the setting of Lucas and Sargent (1981). A state vector

---

<sup>20</sup>See J. Bradford DeLong (1997) and John Taylor (1997).

$x_t \in X$  evolves according to the possibly nonlinear stochastic difference equation

$$x_{t+1} - x_t = f(x_t, t, u_t, v_t, \epsilon_{t+1}) \quad (33)$$

where  $u_t \in U$  is a vector of decisions of private agents,  $v_t \in V$  is a vector of decisions by the government, and  $\epsilon_t \in \mathcal{E}$  is an i.i.d. sequence of random variables with cumulative distribution function  $\Phi$ . A particular example of (33) is

$$x_{t+1} - x_t = \mu(x_t, t, u_t, v_t) + \sigma(x_t, t)\epsilon_{t+1} \quad (34)$$

where  $\Phi$  is Gaussian. Borrowing terms from the corresponding continuous time diffusion specification, we call  $\mu$  the drift and  $\sigma$  the volatility.

Suppose that  $u_t$  and  $v_t$  are governed by the sequences of history-dependent policy functions

$$u_t = h(x^t, t), \quad (35)$$

$$v_t = g(x^t, t), \quad (36)$$

where  $x^t$  denotes the history of  $x_s, s = 0, \dots, t$ . Under the sequences of decision rules (35) and (36), equation (34) becomes

$$x_{t+1} - x_t = \mu(x_t, t, h(x^t, t), g(x^t, t)) + \sigma(x_t, t)\epsilon_{t+1}. \quad (37)$$

This is a nonlinear vector autoregression with stochastic volatility.

Economic theory restricts  $h$  and  $g$ . Private agents' optimum problems and market equilibrium conditions imply a mapping<sup>21</sup>

$$h = T_h(f, g) \quad (38)$$

from the technology and information process  $f$  and the government policy  $g$  to the private sector's equilibrium policy  $h$ . Given  $T_h$ , the normative theory of economic policy would have the government choose  $g$  as the solution of the problem

$$\max_{g, h} E \sum_{t=0}^{\infty} \beta^t W(x_t, u_t, v_t) \quad (39)$$

where  $W$  is a one-period welfare criterion and the optimization is subject to (34) and (38). Notice that the government chooses both  $g$  and  $h$ , although its manipulation of  $h$  is subject to (38). Problem (39) is called a Stackelberg or Ramsey problem.

---

<sup>21</sup>See Stokey (1989) for a description of how households' optimum problems and market clearing are embedded in the mapping (38). Stokey clearly explains why the policies  $h, g$  are history dependent.

Lucas’s (1976) Critique was directed against a faulty econometric policy evaluation procedure that ignores constraint (38). The faulty policy evaluation problem is<sup>22</sup>

$$\max_g E \sum_{t=0}^{\infty} \beta^t W(x_t, u_t, v_t) \quad (40)$$

subject to (34) and  $h = \hat{h}$ , where  $\hat{h}$  is a fixed sequence of decision rules for the private sector. Lucas pointed out first that problem (40) ignores (38) and second that a particular class of models that had been used for  $\hat{h}$  were misspecified because they imputed irrational expectations to private decision makers. Let us express the government’s possibly misspecified econometric model for  $\hat{h}$  through

$$\hat{h} = S(f, g, h), \quad (41)$$

which maps the truth as embodied in the  $f, g, h$  that actually generate the data into the government’s beliefs about private agents’ behavior. The function  $S$  embodies the government’s model specification and also its estimation procedures. See Sargent (1999) for a concrete example of  $S$  within a model of the Phillips curve.

The faulty policy evaluation problem (40) induces

$$g = T_g(f, \hat{h}). \quad (42)$$

The heart of the Lucas critique is that this mapping does not solve the appropriate policy problem (39).

## A.1 Positive implications of imperfect policy making

What outcomes should we expect under the faulty econometric policy evaluation procedure? The answer depends partly on how the government’s econometric estimates  $\hat{h}$  respond to observed outcomes through the function (41). Suppose that the government begins with an initial specification  $\hat{h}_0$  and consider the following iterative process for  $j \geq 1$ :

$$g_j = T_g(f, \hat{h}_{j-1}), \quad (43)$$

$$h_j = T_h(f, g_j), \quad (44)$$

$$\hat{h}_j = S(f, g_j, h_j). \quad (45)$$

In step (43), for fixed  $\hat{h}_{j-1}$ , the government solves the faulty policy problem (40); in step (44) the private sector responds to the government policy  $g_j$ ; in step (45), the government adjusts its econometric model  $\hat{h}_j$  to reflect outcomes under government policy  $g_j$ . We can write the iterative process more compactly as

$$g_j = B(f, g_{j-1}) \quad (46)$$

---

<sup>22</sup>Sargent (1999) calls this a “Phelps problem”.



where  $B(f, g_{j-1}) = T_g(f, S(f, g_{j-1}, T_h(f, g_{j-1})))$ . Eventually, this iterative process might settle down to a fixed point

$$g = B(f, g). \tag{47}$$

In the spirit of Fudenberg and Levine (1993), Fudenberg and Kreps (1995), and Sargent (1999), a *self-confirming equilibrium* is a government policy  $g$  that satisfies (47) and an associated government belief  $\hat{h}$ .

In the following subsections, we first use the iterative scheme (43), (44), (45) to make contact with part of Lucas’s critique. Then we relate the fixed point (47) to the views and practices of Sims (1982, 1999).

## A.2 Adaptation: reconciling two parts of the Lucas critique

Lucas’s (1976) Critique consisted of two parts. The first part of Lucas’s paper summarized empirical evidence for drift in representations like (34), that is, dependence of  $\mu$  on  $t$ , and interpreted it as evidence against particular econometric specifications that had attributed suboptimal forecasts about  $(x, v)$  to private agents. The second part of his paper focused on three concrete examples designed to show how the mapping (38) from  $g$  to  $h$  would influence time series outcomes. Though Lucas didn’t explicitly link the first and second parts, a reader can be forgiven for thinking that he meant to suggest that a substantial part of the drift in  $\mu$  described in the first part of his paper came from drift in private agents’ decision rules that had been induced through mapping (38) by drift in government decision rules.

If we could somehow make a version of the iterative process (43), (44), (45) occur in real time, we get a model of coefficient drift that is consistent with this vision. The literature on least squares learning gets such a real time model by attributing to both private agents and the government a sophisticated kind of adaptive behavior in which the mappings  $T_g, T_h, S$  play key roles. This literature uses recursive versions of least squares learning to deduce drift in  $g$  whose average behavior can eventually be described by the ordinary differential equation<sup>23</sup>

$$\frac{d}{dt}g = B(f, g) - g. \tag{48}$$

In this way it is possible to use the transition dynamics of adaptive systems based on (43), (44), (45) to explain the parameter drift that Lucas emphasized in the first part of his critique. Sargent (1999) and Cho, Williams, and Sargent (2002) pursue

---

<sup>23</sup>See Sargent (1999) and Evans and Honkapohja (2001) for examples and for precise statements of the meanings of ‘average’ and ‘eventually’. Equation (48) embodies the ‘mean dynamics’ of the system. See Cho, Williams, and Sargent (2002) and Sargent (1999). They also describe how ‘escape dynamics’ can be used to perpetuate adaptation.

this line and use it to build models of drifting unemployment-inflation dynamics.<sup>24,25</sup>

### A.3 Another view: asserting a self-confirming equilibrium

Another view takes the data generating mechanism to be the self-confirming equilibrium composed of (47) and (34), unadorned by any transition dynamics based on (43), (44), (45).<sup>26</sup> This view assumes that any adaptation had ended before the sample began. It would either exclude parameter drift or else would interpret it as consistent with a self-confirming equilibrium.<sup>27</sup> Thus, parameter drift would reflect nonlinearities in the law of motion (34) that are accounted for in decision making processes (i.e., hyperparameters would not be drifting). That  $g$  is a fixed point of (47) either excludes government policy regime shifts or requires that they be interpreted as equilibrium government best responses that are embedded in the mapping  $T_g$  in (42) and that are discounted by private agents in the mapping  $T_h$ .

### A.4 Empirical issues

Inspired by theoretical work within the adaptive tradition that permits shifts in policy outside of a self-confirming equilibrium, our earlier paper (Cogley and Sargent (2001)) used a particular nonlinear vector autoregression (37) to compile evidence about how the systematic part of the autoregression,  $x_t + \mu(x^t)$  in (37), has drifted over time. Our specification excluded stochastic volatility (we assumed that  $\sigma(x_t, t) = \bar{\sigma}$ ). We appealed to adaptive models and informally interpreted the patterns of ‘drifting coefficients’ in our nonlinear time series model partly as reflecting shifting behavior

---

<sup>24</sup>As Bray and Kreps (1986) and Kreps (1998) describe, before it attains a self-confirming equilibrium, such an adaptive system embodies irrationality because, while the self-confirming equilibrium is a rational expectations equilibrium, the least squares transition dynamics are not. During the transition, both government and private agents are basing decisions on subjective models that ignore sources of time-dependence in the actual stochastic process that are themselves induced by the transition process. Bray and Kreps (1986) and Kreps (1998) celebrate this departure from rational expectations because they want models of learning *about* a rational expectations equilibrium, not learning *within* a rational expectations equilibrium.

<sup>25</sup>In their Phillips curve example, Kydland and Prescott (1977) explicitly use an example of system (43), (44), (45) and compute its limit to argue informally that inflation would converge to a suboptimal ‘time consistent’ level. Unlike Lucas (1976), Kydland and Prescott’s mapping (45) was  $\hat{h} = h$ . Lucas’s focus was partly to criticize versions of mapping (45) that violated rational expectations, but that was not Kydland and Prescott’s concern.

<sup>26</sup>The literature on least squares learning itself provides substantial support for this perspective by proving almost sure convergence to a self-confirming equilibrium. Sargent (1999) and Cho, Williams, and Sargent (2002) arrest such convergence by putting some forgetting or discounting into least squares.

<sup>27</sup>Sargent and Wallace (1976), Sims (1982), and Sargent (1984) have all expressed versions of this point of view.

rules of the Fed, shifts due to the Fed’s changing preferences or views of the economy.<sup>28</sup>

Sims (1999) and Bernanke and Mihov (1998a, 1998b) analyzed a similar data set in a way that seems compatible with a self-confirming equilibrium within a linear time-invariant structure. They used specializations of the vector time series model (37) that incorporate stochastic volatility but not drift in the systematic part of a linear vector autoregression. Their models can be expressed as

$$x_{t+1} - x_t = Ax_t + \sigma(x_t, t)\epsilon_{t+1}, \quad (49)$$

where we can regard  $x_t$  as including higher order lags of variables and  $A$  is composed of companion submatrices. They compiled evidence that this representation fits post World War II data well and used it to interpret the behavior of the monetary authorities. They found that the systematic part of the vector autoregression  $A$  did not shift over time, but that there was stochastic volatility ( $\sigma(x_t, t) \neq \bar{\sigma}$ ). Thus, they reconciled the data with a linear autoregression in which shocks drawn from time-varying distributions nevertheless feed through the system linearly in a time-invariant way. They reported a lack of evidence for alterations in policy rules (in contrast to the perspective taken for example by Clarida, Gali, and Gertler (2000)).

## A.5 Generalization

In this paper, we fit a model of the form (37) that, permits *both* drifting coefficients *and* stochastic volatility, thereby generalizing both our earlier model and some of the specifications of Bernanke and Mihov and Sims. We use this specification to confront criticisms from Sims (2001) and Stock (2001), both of whom suggested that our earlier results were mainly artifacts of our exclusion of stochastic volatility.

## B A Markov Chain Monte Carlo Algorithm for Simulating the Posterior Density

We use MCMC methods to simulate the posterior density

$$p(\theta^T, Q, \sigma, \beta, H^T | Y^T). \quad (50)$$

By adapting the arguments in section 2.4 of Cogley and Sargent (2001), the posterior can be expressed as

$$p(\theta^T, Q, \sigma, \beta, H^T | Y^T) \propto I(\theta^T)p_L(\theta^T, Q, \sigma, \beta, H^T | Y^T). \quad (51)$$

The function  $I(\theta^T) = \prod_{t=0}^T I(\theta_t)$  is a composite indicator that encodes the VAR stability condition for each date in the sample, and  $p_L(\cdot | Y^T)$  is the posterior that

---

<sup>28</sup>Partly we appealed to adaptive models like ones described by Sims (1988) and Sargent (1999), which emphasize changes in the Fed’s understanding of the structure of the economy.

would result if the nonlinear transition equation (2) were replaced with the linear transition equation (3). As in our earlier paper, we characterize and simulate the unrestricted posterior  $p_L(\cdot | Y^T)$ , and then use rejection sampling to rule out explosive outcomes. This delivers a draw from the nonlinear model,  $p(\cdot | Y^T)$ .

## B.1 Sampling from $p_L(\theta^T, Q, \sigma, \beta, H^T | Y^T)$

We combine the techniques used in Cogley and Sargent (2001) with those of Jacquier, Polson, and Rossi (1994) to construct a Metropolis-within-Gibbs sampler for simulating draws from  $p_L(\cdot | Y^T)$ . The Metropolis-within-Gibbs algorithm consists of 5 steps, one for  $\theta^T$ ,  $Q$ ,  $\beta$ , the elements of  $\sigma$ , and the elements of  $H^T$ . Our prior is that the blocks of parameters are mutually independent, and we assume the marginal prior for each block has a natural conjugate form; details are given above. The first two steps of the algorithm are essentially the same as in our earlier paper,  $\beta$  is treated as a vector of regression parameters, and the elements of  $\sigma$  are treated as inverse-gamma variates. To sample  $H^T$ , we apply a univariate algorithm from Jacquier, et. al. to each element. This is possible because the stochastic volatilities are assumed to be independent.

### B.1.1 VAR parameters, $\theta^T$

We first consider the distribution of VAR parameters conditional on the data and other blocks of parameters. Conditional on  $H^T$  and  $\beta$ , one can calculate the entire sequence of variances  $R_t$ ; we denote this sequence by  $R^T$ . Conditional on  $R^T$  and  $Q$ , the joint posterior density for VAR parameters can be expressed as<sup>29</sup>

$$p_L(\theta^T | Y^T, Q, R^T) = f(\theta_T | Y^T, Q, R^T) \prod_{t=1}^{T-1} f(\theta_t | \theta_{t+1}, Y^t, Q, R_t). \quad (52)$$

The unrestricted model is a linear, conditionally Gaussian state-space model. Assuming a Gaussian prior for  $\theta_0$ , all the conditional densities on the right hand side of (52) are Gaussian. Their means and variances can be computed via a forward and backward recursion.

The forward recursion uses the Kalman filter. Let

$$\begin{aligned} \theta_{t|t} &\equiv E(\theta_t | Y^t, Q, R^T), \\ P_{t|t-1} &\equiv Var(\theta_t | Y^{t-1}, Q, R^T), \\ P_{t|t} &\equiv Var(\theta_t | Y^t, Q, R^T), \end{aligned} \quad (53)$$

represent conditional means and variances going forward in time. These can be

---

<sup>29</sup>The elements of  $\sigma$  are redundant conditional on  $H^T$ .

computed recursively, starting from the prior mean and variance for  $\theta_0$ ,

$$\begin{aligned} K_t &= P_{t|t-1}X_t(X_t'P_{t|t-1}X_t + R_t)^{-1}, \\ \theta_{t|t} &= \theta_{t-1|t-1} + K_t(y_t - X_t'\theta_{t-1|t-1}), \\ P_{t|t-1} &= P_{t-1|t-1} + Q, \\ P_{t|t} &= P_{t|t-1} - K_tX_t'P_{t|t-1}. \end{aligned} \tag{54}$$

At the end of the sample, the forward recursion delivers the mean and variance for  $\theta_T$ , and this pins down the first term in (52),

$$f(\theta_T | Y^T, Q, R^T) = N(\theta_{T|T}, P_{T|T}). \tag{55}$$

The remaining terms in (52) are derived from a backward recursion, which updates conditional means and variances to reflect the additional information about  $\theta_t$  contained in  $\theta_{t+1}$ . Let

$$\begin{aligned} \theta_{t|t+1} &\equiv E(\theta_t | \theta_{t+1}, Y^t, Q, R^T), \\ P_{t|t+1} &\equiv \text{Var}(\theta_t | \theta_{t+1}, Y^t, Q, R^T), \end{aligned} \tag{56}$$

represent updated estimates of the mean and variance. Because  $\theta_t$  is conditionally normal, these are

$$\begin{aligned} \theta_{t|t+1} &= \theta_{t|t} + P_{t|t}P_{t+1|t}^{-1}(\theta_{t+1} - \theta_{t|t}), \\ P_{t|t+1} &= P_{t|t} - P_{t|t}P_{t+1|t}^{-1}P_{t|t}. \end{aligned} \tag{57}$$

The updated estimates determine the mean and variance for remaining elements in (52),

$$f(\theta_t | \theta_{t+1}, Y^T, Q, R^T) = N(\theta_{t|t+1}, P_{t|t+1}). \tag{58}$$

A random trajectory for  $\theta^T$  is generated by iterating backward. The backward recursion starts with a draw of  $\theta_T$  from (55). Then, conditional on its realization,  $\theta_{T-1}$  is drawn from (58),  $\theta_{T-2}$  is drawn conditional on the realization of  $\theta_{T-1}$ , and so on back to the beginning of the sample.

### B.1.2 Innovation Variance for VAR Parameters, $Q$ :

The next step involves the distribution of  $Q$  conditional on the data and other parameter blocks. Conditional on a realization for  $\theta^T$ , the VAR parameter innovations,  $v_t$ , are observable. Furthermore, the other conditioning variables are irrelevant at this stage,

$$p(Q | Y^T, \theta^T, \sigma, \beta, H^T) = p(Q | Y^T, \theta^T). \tag{59}$$

Knowledge of  $\sigma$  is redundant conditional on  $H^T$ , and  $\beta$  and  $H^T$  are irrelevant because  $v_t$  is independent of  $\xi_t$  and  $\eta_{it}$ .

Under the linear transition law,  $v_t$  is *iid* normal. The natural conjugate prior in this case is an inverse-Wishart distribution, with scale parameter  $\bar{Q}$  and degrees of freedom  $T_0$ . Given an inverse-Wishart prior and a normal likelihood, the posterior is inverse-Wishart,

$$p(Q|Y^T, \theta^T) = IW(Q_1^{-1}, T_1), \quad (60)$$

with scale and degree-of-freedom parameters,

$$\begin{aligned} Q_1 &= \bar{Q} + \sum_{t=1}^T v_t v_t', \\ T_1 &= T_0 + T. \end{aligned} \quad (61)$$

### B.1.3 Standard Deviation of Volatility Innovations, $\sigma$

The third step involves the full conditional distribution for  $\sigma$ ,

$$p(\sigma|\theta^T, H^T, \beta, Q, Y^T). \quad (62)$$

Knowledge of  $Q$  is redundant conditional on  $\theta^T$ . The latter conveys information about  $v_t$  and  $\varepsilon_t$ , but both are conditionally independent of the volatility innovations.<sup>30</sup> Thus, conditioning on  $\theta^T$  is also irrelevant.  $\beta$  orthogonalizes  $R_t$  and therefore carries information about  $H_t$ , but this is redundant given direct observations on  $H_t$ . Given a realization for  $H^T$ , one can compute the scaled volatility innovations,  $\sigma_i \eta_{it}$ ,  $i = 1, \dots, 3$ . Because the volatility innovations are mutually independent, we can work with the full conditional density for each. Therefore the density for  $\sigma_1$  simplifies to

$$p(\sigma_1|\sigma_2, \sigma_3, \theta^T, H^T, \beta, Q, Y^T) = p(\sigma_1|h_1^T, Y^T), \quad (63)$$

and similarly for  $\sigma_2^2$  and  $\sigma_3^2$ .

The scaled volatility innovations are *iid* normal with mean zero and variance  $\sigma_i^2$ . Assuming an inverse-gamma prior with scale parameter  $\delta_0$  and  $v_0$  degrees of freedom, the posterior is also inverse gamma,

$$p(\sigma_i^2|h_i^T, Y^T) = IG\left(\frac{v_1}{2}, \frac{\delta_1}{2}\right), \quad (64)$$

where

$$v_1 = v_0 + T, \quad (65)$$

$$\delta_1 = \delta_0 + \sum_{t=1}^T (\Delta \ln h_{it})^2. \quad (66)$$

---

<sup>30</sup>The measurement innovations are informative for  $R_t$ , which depends indirectly on  $\sigma$ , but this information is subsumed in  $H_t$ .

### B.1.4 Covariance Parameters, $\beta$

Next, we consider the distribution of  $\beta$  conditional on the data and other parameters. Knowledge of  $\theta^T$  and  $Y^T$  implies knowledge of  $\varepsilon_t$ , which satisfies

$$B\varepsilon_t = u_t, \quad (67)$$

where  $u_t$  is a vector of orthogonalized residuals with known error variance  $H_t$ . We interpret this as a system of unrelated regressions. The first equation in the system is the identity

$$\varepsilon_{1t} = u_{1t}. \quad (68)$$

The second and third equations can be expressed as transformed regressions,

$$\begin{aligned} (h_{2t}^{-1/2} \varepsilon_{2t}) &= \beta_{21}(-h_{2t}^{-1/2} \varepsilon_{1t}) + (h_{2t}^{-1/2} u_{2t}), \\ (h_{3t}^{-1/2} \varepsilon_{3t}) &= \beta_{31}(-h_{3t}^{-1/2} \varepsilon_{1t}) + \beta_{32}(-h_{3t}^{-1/2} \varepsilon_{2t}) + (h_{3t}^{-1/2} u_{3t}), \end{aligned} \quad (69)$$

with independent standard normal residuals.

Once again, many of the conditioning variables drop out.  $Q$  and  $\sigma$  are redundant conditional on  $\theta^T$  and  $H^T$ , respectively, and  $h_j^T$ ,  $j \neq i$ , are irrelevant because the elements of  $u_t$  are independent. Assuming a normal prior for the regression coefficients in each equation,

$$p(\beta_i) = N(\beta_{i0}, V_{i0}), \quad i = 2, 3, \quad (70)$$

the posterior is also normal,

$$p(\beta_i | Y^T, \theta^T, h_i^T) = N(\beta_{i1}, V_{i1}), \quad i = 2, 3, \quad (71)$$

where

$$V_{i1} = (V_{i0}^{-1} + Z_i' Z_i)^{-1}, \quad (72)$$

$$\beta_{i1} = V_{i1}(V_{i0}^{-1} \beta_{i0} + Z_i' z_i). \quad (73)$$

The variables  $z_i$  and  $Z_i$  refer to the left and right-hand variables, respectively, in the transformed regressions.

### B.1.5 Stochastic Volatilities, $H^T$

The final step involves the conditional distribution of the elements of  $H^T$ . To sample the stochastic volatilities, we apply the univariate algorithm of Jacquier, et. al. (1994) to each element of the orthogonalized VAR residuals,  $u_t$ . The latter are observable

conditional on  $Y^T$ ,  $\theta^T$ , and  $B$ . We can proceed on a univariate basis because the stochastic volatilities are mutually independent.

Jacquier, et. al. adopted a date-by-date blocking scheme and developed the conditional kernel for

$$p(h_{it}|h_{-it}, u_i^T, \sigma_i) = p(h_{it}|h_{it-1}, h_{it+1}, u_i^T, \sigma_i), \quad (74)$$

where  $h_{-it}$  represents the vector of  $h$ 's at all other dates. The simplification follows from the assumption that  $h_{it}$  is Markov. Knowledge of  $Q$  is redundant given  $\theta^T$ , and  $h_j^T$  and  $\sigma_j$ ,  $i \neq j$ , are irrelevant because the stochastic volatilities are independent. By Bayes' theorem, the conditional kernel can be expressed as<sup>31</sup>

$$\begin{aligned} p(h_{it}|h_{it-1}, h_{it+1}, u_i^T, \sigma_i) &\propto p(u_{it}|h_{it})p(h_{it}|h_{it-1})p(h_{it+1}|h_{it}), \\ &\propto h_{it}^{-1.5} \exp\left(-\frac{u_{it}^2}{2h_{it}}\right) \exp\left(\frac{-(\ln h_{it} - \mu_{it})^2}{2\sigma_c^2}\right). \end{aligned} \quad (75)$$

Its form follows from the normal form of the conditional likelihood,  $p(u_{it}|h_{it})$ , and the log-normal form of the log-volatility equation, (11). The parameters  $\mu_{it}$  and  $\sigma_{ic}^2$  are the conditional mean and variance of  $h_{it}$  implied by (11) and knowledge of  $h_{it-1}$  and  $h_{it+1}$ . In the random walk case, they are

$$\begin{aligned} \mu_{it} &= (1/2)(\ln h_{it+1} + \ln h_{it-1}), \\ \sigma_{ic}^2 &= (1/2)\sigma_i^2. \end{aligned} \quad (76)$$

Notice that the normalizing constant is absent from (75). Jacquier, et. al. say the normalizing constant is costly to compute, and they recommend a Metropolis step instead of a Gibbs step. One natural way to proceed is to draw a trial value for  $h_{it}$  from the log-normal density implied by (11), and then use the conditional likelihood  $p(u_{it}|h_{it})$  to compute the acceptance probability. Thus, our proposal density is

$$q(h_{it}) \propto h_{it}^{-1} \exp\left(\frac{-(\ln h_{it} - \mu_{it})^2}{2\sigma_{ic}^2}\right), \quad (77)$$

and the acceptance probability for the  $m$ th draw is

$$\begin{aligned} \alpha_m &= \frac{p(u_{it}|h_{it}^m)q(h_{it}^m)}{q(h_{it}^m)} \frac{q(h_{it}^{m-1})}{p(u_{it}|h_{it}^{m-1})q(h_{it}^{m-1})}, \\ &= \frac{(h_{it}^m)^{-1/2} \exp(-u_{it}^2/2h_{it}^m)}{(h_{it}^{m-1})^{-1/2} \exp(-u_{it}^2/2h_{it}^{m-1})}. \end{aligned} \quad (78)$$

We set  $h_{it}^m = h_{it}^{m-1}$  if the proposal is rejected. The algorithm is applied on a date-by-date basis to each of the elements of  $u_t$ .

---

<sup>31</sup>The formulas are a bit different at the beginning and end of the sample.



## B.2 Rejection Sampling

The Metropolis-within-Gibbs sampler produces draws from the unrestricted posterior,  $p_L(\cdot | Y^T)$ . To sample from the target density,  $p(\cdot | Y^T)$ , we must also enforce the stability condition. We do so by computing the VAR roots associated with  $\theta_t$  at every date and discarding explosive realizations. This can be justified by adapting the arguments in section 2.6 of our earlier paper.

## C Derivation of the State Transition Density

The transition density is defined as

$$p(\theta_{t+1} | \theta_t, Q) = \frac{p(\theta_{t+1}, \theta_t | Q)}{p(\theta_t | Q)}. \quad (79)$$

The numerator can be expressed as

$$p(\theta_{t+1}, \theta_t | Q) = \iint p(\theta^T | Q) d\theta^{t-1} d\theta^{t+2, T}, \quad (80)$$

where  $\theta^{t-1}$  represents the history of  $\theta_s$  up to date  $t-1$  and  $\theta^{t+2, T}$  represents the path from dates  $t+2$  to  $T$ . After substituting from equation (2), this becomes

$$p(\theta_{t+1}, \theta_t | Q) = \frac{1}{m(Q)} \iint \prod_{s=0}^{T-1} I(\theta_{s+1}) f(\theta_{s+1} | \theta_s, Q) d\theta^{t-1} d\theta^{t+2, T}, \quad (81)$$

where  $m(Q) = \int I(\theta^T) f(\theta^T | Q) d\theta^T$  is the normalizing constant for (2). The integrand can be expanded as

$$\begin{aligned} \prod_{s=0}^{T-1} I(\theta_{s+1}) f(\theta_{s+1} | \theta_s, Q) &= \prod_{s=0}^{t-1} I(\theta_{s+1}) f(\theta_{s+1} | \theta_s, Q) \\ &\quad \times I(\theta_{t+1}) f(\theta_{t+1} | \theta_t, Q) \\ &\quad \times \prod_{s=t+1}^{T-1} I(\theta_{s+1}) f(\theta_{s+1} | \theta_s, Q). \end{aligned} \quad (82)$$

It follows that

$$\begin{aligned} p(\theta_{t+1}, \theta_t | Q) &= \frac{1}{m(Q)} \int \prod_{s=0}^{t-1} I(\theta_{s+1}) f(\theta_{s+1} | \theta_s, Q) d\theta^{t-1} \\ &\quad \times I(\theta_{t+1}) f(\theta_{t+1} | \theta_t, Q) \\ &\quad \times \int \prod_{s=t+1}^{T-1} I(\theta_{s+1}) f(\theta_{s+1} | \theta_s, Q) d\theta^{t+2, T}. \end{aligned} \quad (83)$$

The marginal density for  $\theta_t$  can be expressed as

$$\begin{aligned}
p(\theta_t|Q) &= \int p(\theta_{t+1}, \theta_t|Q) d\theta_{t+1}, \\
&= \frac{1}{m(Q)} \int \prod_{s=0}^{t-1} I(\theta_{s+1}) f(\theta_{s+1}|\theta_s, Q) d\theta^{t-1} \\
&\quad \times \int \prod_{s=t}^{T-1} I(\theta_{s+1}) f(\theta_{s+1}|\theta_s, Q) d\theta^{t+1, T}.
\end{aligned} \tag{84}$$

The transition density is the ratio of (83) to (84),

$$\begin{aligned}
p(\theta_{t+1}|\theta_t, Q) &= \frac{I(\theta_{t+1})f(\theta_{t+1}|\theta_t, Q) \int \prod_{s=t+1}^{T-1} I(\theta_{s+1})f(\theta_{s+1}|\theta_s, Q)d\theta^{t+2, T}}{\int \prod_{s=t}^{T-1} I(\theta_{s+1})f(\theta_{s+1}|\theta_s, Q)d\theta^{t+1, T}}, \\
&= \frac{I(\theta_{t+1})f(\theta_{t+1}|\theta_t, Q) \int I(\theta^{t+2, T})f(\theta^{t+2, T}|\theta_{t+1}, Q)d\theta^{t+2, T}}{\int I(\theta_{t+1})f(\theta_{t+1}|\theta_t, Q) [\int I(\theta^{t+2, T})f(\theta^{t+2, T}|\theta_{t+1}, Q)d\theta^{t+2, T}] d\theta_{t+1}}.
\end{aligned} \tag{85}$$

The integral in the numerator is the expectation of  $I(\theta^{t+2, T})$  with respect to the conditional density  $f(\theta^{t+2, T}|\theta_{t+1}, Q)$ . This represents the probability that random walk trajectories emanating from  $\theta_{t+1}$  will remain in the nonexplosive region from date  $t+2$  through date  $T$ . In the text, this term is denoted  $\pi(\theta_{t+1}, Q)$ . Hence the transition density is

$$p(\theta_{t+1}|\theta_t, Q) = \frac{I(\theta_{t+1})f(\theta_{t+1}|\theta_t, Q)\pi(\theta_{t+1}, Q)}{\int I(\theta_{t+1})f(\theta_{t+1}|\theta_t, Q)\pi(\theta_{t+1}, Q)d\theta_{t+1}}. \tag{86}$$

## References

Aguilar, Omar and Mike West, 2001, "Bayesian Dynamic Factor Models and Portfolio Allocation," *Journal of Business and Economic Statistics*.

Anderson, Evan, Lars Peter Hansen, and Thomas J. Sargent, 2000, "Robustness, Detection, and the Price of Risk," Mimeo, Department of Economics, Stanford University.

Andrews, Donald W.K., 1993, "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica* 61, pp. 821-856.

Benati, Luca, 2001, "Investigating Inflation Dynamics Across Monetary Regimes: Taking the Lucas Critique Seriously," Bank of England working paper.

Bernanke, Ben S. and Ilian Mihov, 1998a, "The Liquidity Effect and Long-Run Neutrality." In *Carnegie-Rochester Conference Series on Public Policy*, 49, Bennett T. McCallum and Charles I. Plosser, eds. (Amsterdam: North Holland), pp. 149-194.

- and —, 1998b, “Measuring Monetary Policy.” *Quarterly Journal of Economics*, 113, August, pp. 869-902.
- Boivin, Jean, 1999, “Revisiting the Evidence on the Stability of Monetary VARs,” unpublished manuscript, Graduate School of Business, Columbia University.
- Bray, Margaret M. and David Kreps, 1986, “Rational Learning and Rational Expectations,” in W. Heller, R. Starr, and D. Starrett, eds., *Essays in Honor of Kenneth J. Arrow* (Cambridge University Press: Cambridge, UK).
- Cho, In Koo, Noah Williams, and Thomas J. Sargent, 2002, “Escaping Nash Inflation,” *Review of Economic Studies*, Vol. 69, January, pp. 1–40.
- Clarida, Richard, Jordi Gali, and Mark Gertler, 2000, “Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory,” *Quarterly Journal of Economics* 115(1), pp. 147-180.
- Cogley, Timothy and Thomas J. Sargent, 2001, “Evolving Post World War II U.S. Inflation Dynamics,” *NBER Macroeconomics Annual* 16, pp. 331-373.
- DeLong, J. Bradford, 1997, “America’s Only Peacetime Inflation: the 1970’s,” in Christina Romer and David Romer (eds.), *Reducing Inflation*. NBER Studies in Business Cycles, Volume 30.
- Evans, George W. and Seppo Honkapohja, 2001, *Learning and Expectations in Macroeconomics* (Princeton University Press: Princeton, New Jersey).
- Fudenberg, Drew and David K. Levine, 1993, “Self-Confirming Equilibrium,” *Econometrica* 61, pp.523-545.
- and David M. Kreps, 1995, “Learning in Extensive Games, I: Self-Confirming Equilibria,” *Games and Economic Behavior* 8, pp. 20-55.
- Hansen, Bruce E., 1992, “Testing For Parameter Instability in Linear Models,” *Journal of Policy Modeling* 14, pp. 517-533.
- Ireland, Peter, 1999, “Does the Time-Consistency Problem Explain the Behavior of Inflation in the United States?” *Journal of Monetary Economics* 44(2), pp. 279-292.
- Jacquier, Eric, Nicholas G. Polson, and Peter Rossi, 1994, “Bayesian Analysis of Stochastic Volatility Models,” *Journal of Business and Economic Statistics* 12, pp. 371-418.
- , —, and —, 1999, “Stochastic Volatility: Univariate and Multivariate Extensions,” unpublished manuscript, Finance Department, Boston College and Graduate School of Business, University of Chicago.
- Kim, Chang-Jin and Charles R. Nelson, 1999a, “Has The U.S. Economy Become More Stable? A Bayesian Approach Based on a Markov Switching Model of the Business Cycle,” *Review of Economics and Statistics* 81(4), 608-661.
- Kreps, David, 1998, “Anticipated Utility and Dynamic Choice,” Mimeo, 1997 Schwartz Lecture, Northwestern University.

Kydland, Finn and Edward C. Prescott, 1977, "Rules Rather than Discretion: the Inconsistency of Optimal Plans," *Journal of Political Economy* 85, pp. 473-491.

Leeper, Eric and Tao Zha, 2001a, "Empirical Analysis of Policy Interventions," Mimeo, Department of Economics, Indiana University and Research Department, Federal Reserve Bank of Atlanta.

— and —, 2001b, "Toward a Theory of Modest Policy Interventions," Mimeo, Department of Economics, Indiana University and Research Department, Federal Reserve Bank of Atlanta.

Lucas, Robert E., Jr., 1976, "Econometric Policy Evaluation: A Critique," in *The Phillips Curve and Labor Markets*, edited by Karl Brunner and Alan Meltzer, Carnegie-Rochester Series on Public Policy, vol. 1.

— and Thomas J. Sargent, 1981, "Introduction" in Robert E. Lucas and Thomas J. Sargent (eds.) *Rational Expectations and Econometric Practice* (Minneapolis: University of Minnesota Press).

McCallum, Bennett T., 1999, "Issues in the Design of Monetary Policy Rules" in Taylor, John B. and Michael Woodford, eds., *Handbook of Macroeconomics* vol. 1C (Amsterdam: Elsevier Science).

McConnell, Margaret and Gabriel Perez Quiros, 2000, "Output Fluctuations in the United States: What Has Changed Since the Early 1980s?" *American Economic Review* 90(5), 1464-1476.

Nyblom, Jukka, 1989, "Testing for the Constancy of Parameters Over Time," *Journal of the American Statistical Association* 84, pp. 223-230.

Parkin, Michael, 1993, "Inflation in North America," in *Price Stabilization in the 1990s*, edited by Kumiharo Shigehara.

Pitt, Mark and Neil Shepard, 1999, "Time-Varying Covariances: A Factor Stochastic Volatility Approach," in *Bayesian Statistics 6*, J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith, eds., (Oxford University Press: Oxford).

Romer, Christina D. and David H. Romer, 2002, "The Evolution of Economic Understanding and Postwar Stabilization Policy," forthcoming in the 2002 Jackson Hole conference volume, Federal Reserve Bank of Kansas City.

Rudebusch, Glenn D. and Lars E.O. Svensson, 1999, "Policy Rules for Inflation Targeting," in *Monetary Policy Rules*, edited by John B. Taylor, NBER Conference Report (University of Chicago Press: Chicago, Illinois).

Samuelson, Paul A., and Robert M. Solow, 1960. "Analytical Aspects of Anti-Inflation Policy," *American Economic Review*, Vol 50, May, pp. 177-184.

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

— 1984, "Autoregressions, Expectations, and Advice," *American Economic Review, Papers and Proceedings* 74, pp. 408-415.

— and Neil Wallace, 1976, “Rational Expectations and the Theory of Economic Policy,” *Journal of Monetary Economics* 2, pp. 169-183.

Sims, Christopher A., 1980, “Comparison of Interwar and Postwar Business Cycles: Monetarism Reconsidered”, *American Economic Review*, pp. 250-257.

—, 1982, “Policy Analysis with Econometric Models,” *Brookings Papers on Economic Activity*, Vol. 1, pp. 107-152.

—, 1988, “Projecting Policy Effects with Statistical Models,” *Revista de Analisis Economico* 3, pp. 3-20.

—, 1999. “Drifts and Breaks in Monetary Policy,” mimeo, Princeton University.

—, 2001, “Comment on Sargent and Cogley’s ‘Evolving Post World War II U.S. Inflation Dynamics,’” *NBER Macroeconomics Annual* 16, pp 373-379.

Sims, Christopher A. and Tao Zha, 1999, “Error Bands for Impulse Responses,” *Econometrica* 67, pp. 1113-1155.

Stock, James H., 2001, “Discussion of Cogley and Sargent ‘Evolving Post World War II U.S. Inflation Dynamics,’” *NBER Macroeconomics Annual* 16, pp. 379-387.

Stokey, Nancy L., 1989, “Reputation and Time Consistency,” *American Economic Review, Papers and Proceedings* 79, pp. 134-139.

Taylor, John B., (1993), “Discretion versus Policy Rules in Practice,” in *Carnegie-Rochester Conference Series on Public Policy*, Vol. 39, December, pp. 195–214.

— 1997, “Comment on America’s Only Peacetime Inflation: the 1970’s,” in Christina Romer and David Romer (eds.), *Reducing Inflation*. NBER Studies in Business Cycles, Volume 30.

Whittle, Peter, 1953, “The Analysis of Multiple Stationary Time Series,” *Journal of the Royal Statistical Society, Series B*, vol. 15, pp. 125-139.