Available online at www.sciencedirect.com
SCIENCE DIRECT®

Review of Economic Dynamics 8 (2005) 262–302

Review of Economic Dynamics

www.elsevier.com/locate/red

Drifts and volatilities: monetary policies and outcomes in the post WWII US

Timothy Cogley [a], Thomas J. Sargent [b,c,*]

[a] *University of California, Davis, Davis, CA 95616, USA*
[b] *New York University, New York, NY 10003, USA*
[c] *Hoover Institution, Stanford, CA 94305, USA*

Received 1 June 2004

Available online 23 February 2005

Abstract

For a VAR with drifting coefficients and stochastic volatilities, we present posterior densities for several objects that are pertinent 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 US 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 hypothesis of time-invariance of autoregressive coefficients of VARs against the alternative of time-varying coefficients. Except for one, we find that those tests have low power against the form of time variation captured by our model.

* Corresponding author at: Department of Economics, New York University, 269 Mercer Street, 7th Floor, New York, NY 10003, USA.

E-mail address: thomas.sargent@nyu.edu (T.J. Sargent).

1094-2025/$ – see front matter
doi:10.1016/j.red.2004.10.009

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 US 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 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. We find that those tests have low power against our particular model of drifting coefficients, except for one test. And that 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 organizes evidence within a formal statistical model. 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 US 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 selecting interesting anecdotes and passages from government reports, asserts that it was bad policy. Their story is that during the 1950s and early 1960s, the Fed understood a 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 (1960) promise of an exploitable trade-off between inflation and unemployment; and that under Volcker's leadership, the Fed came to its senses, again accepted the natural rate hypothesis, and used monetary policy to arrest inflation.

Aspects of this "Berkeley view"[1] receive backing from statistical work by Clarida et al. (2000) and Taylor (1993), who fit monetary policy rules for subperiods that they choose to

[1] Sargent (2002) summarizes and evaluates the Berkeley view.

isolate differences between the Burns and the Volcker–Greenspan eras. They find evidence for a systematic difference of monetary policies across the two eras, a difference that in Clarida et al.'s 'new-neoclassical-synthesis' macroeconomic model would lead to better inflation-unemployment outcomes under the Volcker–Greenspan policy.

But Taylor's and Clarida et al.'s interpretation of the data has been disputed by Sims (1980, 1999) and Bernanke and Mihov (1998a, 1998b). They presented evidence that the US data do not prompt rejection of an hypothesis of 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 post-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. In particular, as we shall see, inflation persistence also plays an important role in lending relevance to instruments for estimating monetary policy rules.[2] 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 1970s, then fell during the 1980s and 1990s.

1.4. Drifting coefficients and the Lucas's Critique

Drifting coefficients have been an important piece of unfinished business within macroeconomic theory since Lucas (1976) emphasized them in the first half of his 1976 Critique, but then ignored them in the second half.[3] Sargent (1999) reevaluated how drifting coefficients bear on the theory of economic policy in the context of recent ideas about self-confirming equilibria. A self-confirming equilibrium is a subtle extension of the concept of a rational expectations equilibrium that allows different decision makers to have models that differ, but only about the probabilities of events off an equilibrium path. In particular, in a self-confirming equilibrium within a macroeconomic context, a government's model correctly describes conditional probabilities of events that occur infinitely often on the equilibrium path, but it can be wrong about off-equilibrium path occurrences that are important because they enter the thought process by which the government designs its policy. The idea that the observations conform strictly to a self-confirming equilibrium can bolster the time-invariance view of the data taken by Sims and Bernanke and Mihov. However, Sargent (1999) and Cho et al. (2002) show how a slight modification of

[2] Sims and Zha (2004, Appendix A) argue that the identification of monetary policy rules is fragile in models that contain both a forward looking monetary policy rule something like an IS curve. See Sargent (1973) for a discussion of some closely related issues in estimating the 'Fisher equation.'

[3] See Sargent (1999) for more about this interpretation of the two halves of Lucas's (1976) paper.

a self-confirming equilibrium, attained by attributing adaptive behavior to the government, produces a macroeconomic model whose tendencies toward a self-confirming equilibrium are interrupted by recurrent escapes that generate nonlinearities in the data that can show up as drifting coefficients.

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.

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 A a detailed characterization of the priors and posterior for our model, and Appendix B describes a Markov Chain Monte Carlo algorithm that we use to approximate the posterior density. Section 3 reports our results, and Section 4 concludes.

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, but assumed that the VAR innovation variance was constant. Thus, our measurement equation was

$$y_t = X_t'\theta_t + \varepsilon_t, \tag{1}$$

where y_t is a vector of endogenous variables, X_t includes a constant plus lags of y_t, and θ_t is a vector of VAR parameters. We assumed that the residuals, ε_t, were conditionally normal with mean zero and constant covariance matrix R.

We assumed that the VAR parameters evolved as driftless random walks subject to reflecting barriers. Let

$$\theta^T = \left[\theta_1', \ldots, \theta_T'\right]' \tag{2}$$

represent the history of VAR parameters from dates 1 to T. The driftless random walk component is represented by a joint prior,

$$f\left(\theta^T, Q\right) = f\left(\theta^T \mid Q\right) f(Q) = f(Q) \prod_{s=0}^{T-1} f(\theta_{s+1} \mid \theta_s, Q), \tag{3}$$

where

$$f(\theta_{t+1} \mid \theta_t, Q) \sim N(\theta_t, Q). \tag{4}$$

Thus, apart from the reflecting barrier, θ_t evolves as

$$\theta_t = \theta_{t-1} + v_t. \tag{5}$$

The innovation v_t is normal with mean zero and variance Q, and we allowed for correlation between the state and measurement innovations, $\mathrm{cov}(v_t, \varepsilon_t) = C$. The marginal prior $f(Q)$ makes Q an inverse-Wishart variate.

The reflecting barrier was encoded in an indicator function, $I(\theta^T) = \prod_{s=1}^{T} I(\theta_s)$. The function $I(\theta_s)$ takes a value of 0 when the roots of the associated VAR polynomial are inside the unit circle, and it is equal to 1 otherwise. This restriction truncates and renormalizes the random walk prior,

$$p\big(\theta^T, Q\big) \propto I\big(\theta^T\big) f\big(\theta^T, Q\big). \tag{6}$$

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. When 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.[4]

In Appendix A, we derive a number of relations between the restricted and unrestricted priors. Among other things, the restricted prior for $\theta^T \mid Q$ can be expressed as

$$p\big(\theta^T \mid Q\big) = \frac{I(\theta^T) f(\theta^T \mid Q)}{m_\theta(Q)}, \tag{7}$$

the marginal prior for Q becomes

$$p(Q) = \frac{m_\theta(Q) f(Q)}{m_Q}, \tag{8}$$

and the transition density is

$$p(\theta_{t+1} \mid \theta_t, Q) \propto I(\theta_{t+1}) f(\theta_{t+1} \mid \theta_t, Q) \pi(\theta_{t+1}, Q). \tag{9}$$

The terms $m_\theta(Q)$ and m_Q are normalizing constants and are defined in Appendix A.[5]

In (7), the stability condition truncates and renormalizes $f(\theta^T \mid Q)$ to eliminate explosive θ's. In (8), the marginal prior $f(Q)$ is re-weighted by $m_\theta(Q)$, the probability of an explosive draw from $f(\theta^T \mid Q)$. This diminishes the probability of Q-values that are likely to generate explosive θ's. Since large values of Q make explosive draws more likely, this

[4] 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.

[5] These expressions supersede those 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.

shifts the prior probability toward smaller values of Q. In other words, relative to $f(Q)$, $p(Q)$ is tilted in the direction of less time variation in θ. Finally, in (9), $f(\theta_{t+1} \mid \theta_t, Q)$ is truncated and re-weighted by $\pi(\theta_{t+1}, Q)$. The latter term represents the probability that random walk paths emanating from θ_{t+1} will remain in the nonexplosive region going forward in time. Thus, the restricted transition density censors explosive draws from $f(\theta_{t+1} \mid \theta_t, Q)$ and down-weights those likely to become explosive.[6]

2.1. *Sims's and Stock's criticisms*

Sims (2001) and Stock (2001) said that we exaggerated the time variation in θ_t. One issue arose because Cogley and Sargent (2001) reported results based on filtered estimates. Sims pointed out that there is transient variation in filtered estimates even in time-invariant systems. Therefore, in this paper, we report results based on smoothed estimates of θ.

More importantly, Sims and Stock questioned our assumption that R is constant. They pointed to evidence developed by Bernanke and Mihov (1998a, 1998b), 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 1979 and 1982. Kim and Nelson and McConnell and Perez Quiros studied the growing stability of the US economy, which they characterized in terms of a large decline in VAR innovation variances after the mid-1980s. Although the cause of this decline is a subject of debate, 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 θ. 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 θ and varying R.

If the world were characterized by constant θ and drifting R, and we fit an approximating model with constant R and drifting θ, it is possible that our estimates of θ would drift to compensate for misspecification of R, thus exaggerating the time variation in θ. 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 θ and drifting R would attribute the high inflation variance of the 1970s to an increase in innovation variances, while a model with drifting θ 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 could be an artifact of model misspecification.

[6] 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 Eq. (9). What matters is the relative likelihood of future instability, not the absolute likelihood.

2.2. Strategy for sorting out the issues

Although it is possible that the coefficients and the volatilities both vary, 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 θ, and also to conduct simulations of the power of the Andrews–Bernanke–Mihov test. Their null hypothesis, which they were unable to reject, was that θ 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 activate both elements, we retain much of the specification described above, but now we assume that the VAR innovations can be expressed as

$$\varepsilon_t = R_t^{1/2}\xi_t, \tag{10}$$

where ξ_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,

$$E(\xi_t v_s) = 0 \quad \text{for all } t, s. \tag{11}$$

To model drifting variances, we adopt a multivariate version of the stochastic volatility model of Jacquier et al. (1994).[7] In particular, we assume that R_t can be expressed as

$$R_t = B^{-1}H_t B^{-1\prime}, \tag{12}$$

where H_t is diagonal and B is lower triangular:

$$H_t = \begin{pmatrix} h_{1t} & 0 & 0 \\ 0 & h_{2t} & 0 \\ 0 & 0 & h_{3t} \end{pmatrix}, \tag{13}$$

$$B = \begin{pmatrix} 1 & 0 & 0 \\ \beta_{21} & 1 & 0 \\ \beta_{31} & \beta_{32} & 1 \end{pmatrix}. \tag{14}$$

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

$$\ln h_{it} = \ln h_{it-1} + \sigma_i \eta_{it}. \tag{15}$$

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 US economy. The volatility innovations, η_{it}, are standard normal random variables that are independent of one another and of the other shocks in the model, ξ_t and v_t. The volatility innovations are each scaled by a free parameter σ_i that determines their magnitude. The factorization in (12) and log specification in (15) guarantee that R_t is positive definite. The free parameters

[7] This formulation is closely related to the multi-factor stochastic volatility models of Aguilar and West (2000), Jacquier et al. (1999), and Pitt and Shepard (1999).

in B allow for correlation among the elements of ε_t. The matrix B orthogonalizes ε_t, but it is not an identification scheme.

This specification departs from a large literature that uses GARCH models of time-varying volatilities. Our preference for stochastic volatility is mostly a matter of convenience. If conditional mean and variance innovations were related to one another, as they are in GARCH models, the conditional submodels that we exploit in the simulation would not decompose as nicely, and conditional mean and variance parameters could not be assigned to separate blocks. Our stochastic volatility model assumes that η_{it} and ξ_t are independent, and this greatly simplifies the algorithm for simulating the posterior distribution.[8]

The specification also 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 occur only once. Our specification allows recurrent permanent shifts.

We use Markov Chain Monte Carlo (MCMC) methods to simulate the posterior density. Let

$$Y^T = [y_1', \dots, y_T']' \tag{16}$$

and

$$H^T = \begin{bmatrix} h_{11} & h_{21} & h_{31} \\ h_{12} & h_{22} & h_{32} \\ \dots & \dots & \dots \\ h_{1T} & h_{2T} & h_{3T} \end{bmatrix} \tag{17}$$

represent the history of data 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,

$$p\big(\theta^T, Q, \sigma, \beta, H^T \mid Y^T\big), \tag{18}$$

summarizes beliefs about the model's free parameters, conditional on priors and the history of observations, Y^T. Appendix B describes details of the algorithm that we use to simulate this posterior.

[8] These comments will make more sense for readers who work through Appendix B, which provides details on the simulation algorithm.

3. Empirical results

3.1. Data

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 seasonally adjusted. The CPI is point sampled in the third month of each quarter, and quarterly inflation is measured as the log difference of these values. Unemployment, on the other hand, is measured as the quarterly average of monthly rates. The Treasury bill data are the average of daily rates in the first month of each quarter. The sample spans the period 1948.Q1 to 2000.Q4, and we work with VAR(2) representations for nominal interest, quarterly 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,

$$\begin{aligned} &f(\theta_0, h_{10}, h_{20}, h_{30}, Q, \beta, \sigma_1, \sigma_2, \sigma_3) \\ &\quad = f(\theta_0) f(h_{10}) f(h_{20}) f(h_{30}) f(Q) f(\beta) f(\sigma_1) f(\sigma_2) p(\sigma_3). \end{aligned} \tag{19}$$

Our prior for θ_0 is a truncated Gaussian density,

$$p(\theta_0) \propto I(\theta_0) f(\theta_0) = I(\theta_0) N(\bar{\theta}, \bar{P}). \tag{20}$$

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 θ_0 is only weakly restricted.

The matrix Q is a key parameter because it governs the rate of drift in θ. We adopt an informative prior for Q, but set its parameters to maximize the influence of sample information. Our prior for Q is inverse-Wishart,

$$f(Q) = IW(\bar{Q}^{-1}, T_0), \tag{21}$$

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

$$T_0 = \dim(\theta_t) + 1. \tag{22}$$

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

$$\bar{Q} = \gamma^2 \bar{P} \tag{23}$$

and set $\gamma^2 = 3.5\text{e–}04$. This makes $\bar{Q}$ comparable to the value used in Cogley and Sargent (2001).[9]

[9] An earlier draft experimented with alternative values of γ that push $\bar{Q}$ toward zero, i.e. in the direction of less variation in θ. We found only minor sensitivity to changes in γ.

This is a 'business-as-usual' prior, in the sense of Leeper and Zha (2001a, 2001b) because we set $\bar{Q}$ to imply little variation in θ. Had we calibrated $Q = \bar{Q}$, or set T_0 so that a substantial weight was put on values in its neighborhood, drift in posterior estimates of θ would have been negligible. But because it has minimal degrees of freedom, the prior distribution is *weak*. With $T_0 = \dim(\theta_t) + 1$, (21) is proper (i.e., it integrates to 1), but it does not have finite moments. Thus, while the matrix $\bar{Q}$ is conservative relative to our vision of the world, the weakness of the initial prior means that it does not tie us down to that conservative vision as evidence accrues.[10,11]

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

$$f(\ln h_{i0}) = N\big(\ln \bar{h}_i, 10\big), \tag{24}$$

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 β is normal with a large variance,

$$f(\beta) = N(0, 10000 \cdot I_3). \tag{25}$$

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

$$f\big(\sigma_i^2\big) = IG\left(\frac{0.01^2}{2}, \frac{1}{2}\right). \tag{26}$$

This specification is also designed to put a heavy weight on sample information, for (26) also does not possess finite moments.

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, yielding a sample of 5000 draws from the posterior density, which economizes on storage space and reduces autocorrelation across draws, at the cost of increasing the variance of ensemble averages from the simulation. The estimates reported below are computed from averages of this sample.

3.4. The posterior distribution for Q

We begin with evidence on the rate of drift in θ, as summarized by the posterior estimate of Q. Recall that Q is the innovation variance in the unrestricted transition equation for VAR parameters. Large values mean rapid movements in θ, smaller values imply a slower rate of drift, and $Q = 0$ represents a time-invariant model.

---

10 Primiceri (2003a) adopts the same $\bar{Q}$ but doubles T_0, thus increasing the weight of the conservative prior.

11 The vagueness of the prior on Q explains our reluctance to evaluate competing models by calculating Bayes's factors. As Lindley (1957) and others have shown, strange outcomes can occur when a prior is so diffuse.

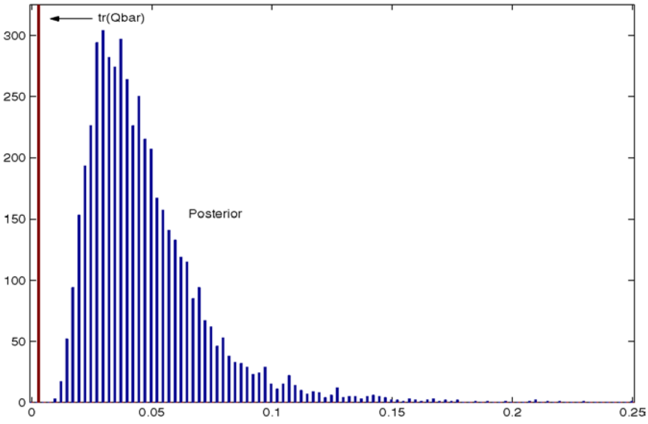

Fig. 1. $\mathrm{tr}(\bar{Q})$ and the posterior distribution for $\mathrm{tr}(Q)$.

First we illustrate how the data alter priors on Q. Figure 1 portrays the posterior histogram for the trace of Q along with the trace of the prior value $\bar{Q}$. The posterior sample is shifted to the right of $\mathrm{tr}(\bar{Q})$; indeed, the smallest value in posterior sample is 0.01, which is 3 times larger than $\mathrm{tr}(\bar{Q})$. Thus the sample points toward greater time variation in θ.

Figure 2 illustrates what this means for θ^T. The top panel depicts the prior mean of θ^T, and the bottom panel portrays the posterior mean. Notice that these estimates integrate across Q in prior and posterior samples, respectively. In the top panel, θ wiggles slightly, but the degree of time variation is barely perceptible. More time variation is evident in the bottom panel, though it is concentrated in particular elements of θ.

Next we address two other questions about the posterior for Q, namely, whether the results are sensitive to the VAR ordering and how the stability prior influences the rate of drift in θ. 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 how 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 determine whether there exists an ordering that mutes evidence for drift in θ, as in Sims (1980). As Table 1 shows, 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 remainder of the paper focuses on the ordering that minimizes the rate of drift in θ, namely $[i_t, u_t, \pi_t]'$. This is conservative from our perspective, but results for the other orderings are similar.

The second question concerns how the stability prior influences drift in θ. One might conjecture that the stability constraint amplifies evidence for drift in θ by pushing the system away from the unit root boundary, forcing the model to fit inflation persistence via

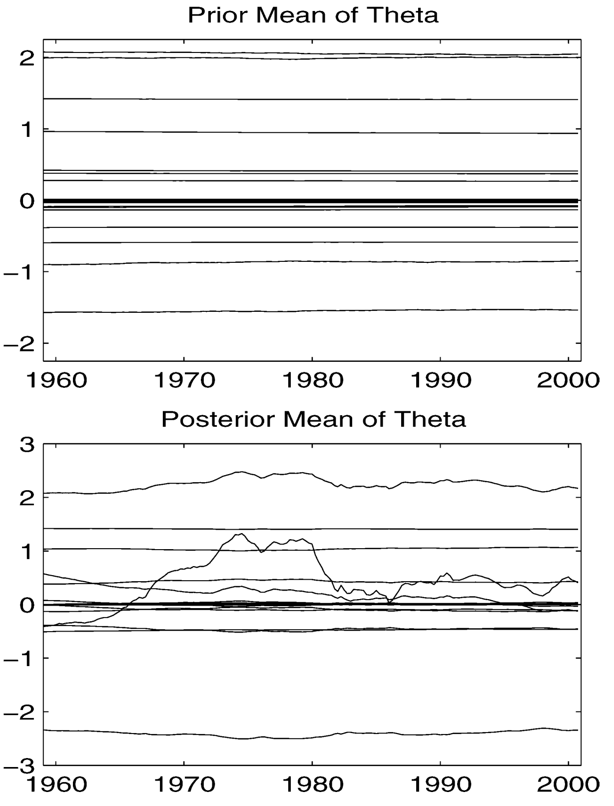

Fig. 2. Prior and posterior mean of θ^T.

Table 1
Posterior mean of Q

VAR orderings	Stability imposed		Stability not imposed	
	tr(Q)	max(λ)	tr(Q)	max(λ)
i, π, u	0.055	0.025	0.056	0.027
i, u, π	0.047	0.023	0.059	0.031
π, i, u	0.064	0.031	0.082	0.044
π, u, i	0.062	0.031	0.088	0.051
u, i, π	0.057	0.026	0.051	0.028
u, π, i	0.055	0.024	0.072	0.035

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

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 θ.

The next table explores the structure of drift in θ, focusing on the minimum-Q ordering $[i, u, \pi]'$. Sargent (1999) learning model predicts that reduced form parameters should drift

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 nth component (the nth 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 $[i, u, \pi]'$.

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).

The table confirms that drift in θ 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, there are only two or three drifting components in θ that manifest themselves various ways.

3.5. The evolution of R_t

Figure 3 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 between pairs of innovations. 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, the “Great Moderation” is evident in the innovation variance for unemployment. The standard deviation dropped by roughly 40 percent in the early 1980s, 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. Our estimates suggest that a comparable decline in variance occurred 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 is likely to continue is an open question.

In addition, there were spikes in the innovation variances for nominal interest and inflation between 1979 and 1981, a pattern that resembles the estimates of Bernanke and

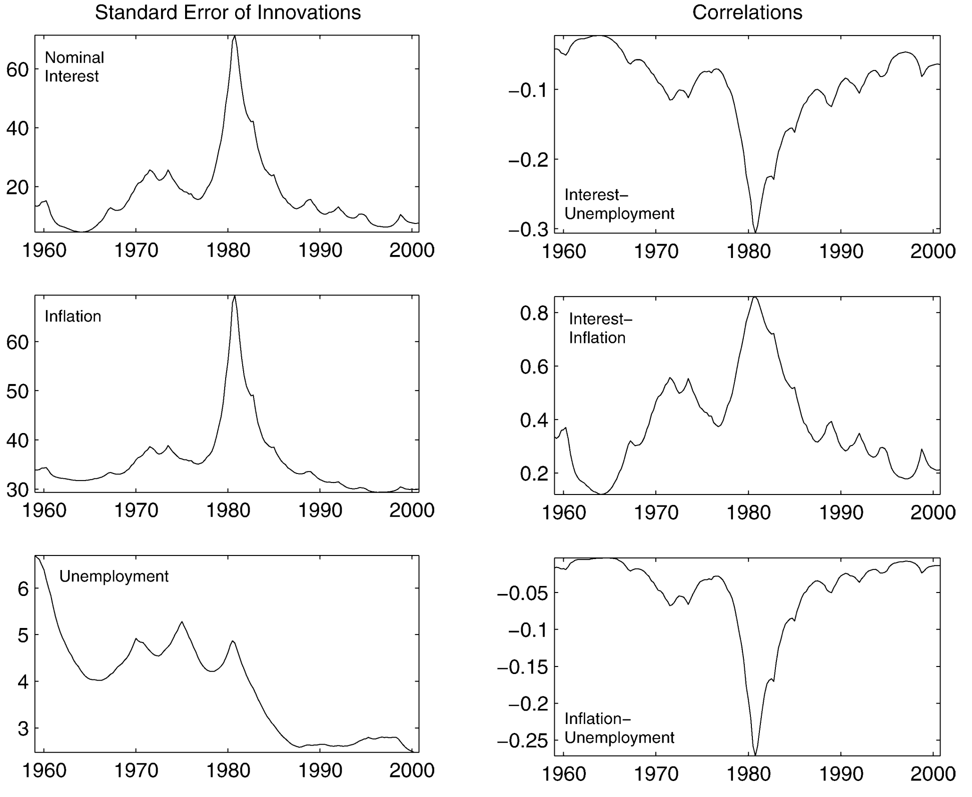

Fig. 3. Posterior mean of R_t.

Mihov. The two variances fell sharply after 1981, however, and reverted within a few years to levels achieved in the 1960s.

The right-hand column shows 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 taken early in the quarter. If the Fed's policy rule depends only on lagged information, 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.

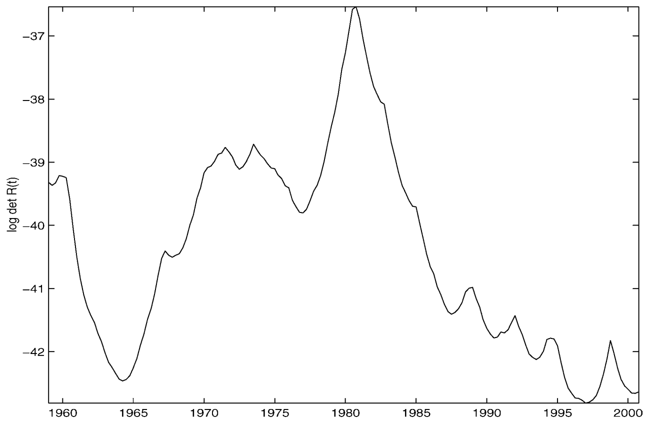

Fig. 4. Total prediction variance.

The signs of the correlations in Fig. 3 are inconsistent with this interpretation. If nominal interest innovations were policy shocks, conventional wisdom suggests that 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. Some information must be missing.[12]

Finally, Fig. 4 reports $\log|E(R_{t|T})|$. Following Whittle (1953), we interpret this as a measure of the total uncertainty entering the system at each date. The smoothed estimates shown here resemble 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 occurred in two steps, one between 1964 and 1972 and the other between 1977 and 1981. Most of the subsequent decrease occurred in the mid-1980s, during the latter part of Volcker's term. This picture suggests that the growing stability documented in the literature on the "Great Moderation" may actually reflect a return to stability. It is interesting to note how short-lived the earlier episode proved to be.

3.6. The evolution of θ_t

Variation in R is clearly an interesting and important feature of the data, but does it alter the patterns of drift in θ 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 contributed to the rise and fall of inflation in the 1970s and 1980s.

[12] 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.

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 we estimate from local linear approximations to mean inflation and unemployment, respectively, evaluated at the posterior mean, $E(\theta_{t|T})$. Write (1) in companion form as

$$z_t = \mu_{t|T} + A_{t|T} z_{t-1} + u_t, \tag{27}$$

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, we approximate mean inflation at t by

$$\bar{\pi}_t = s_\pi (I - A_{t|T})^{-1} \mu_{t|T}, \tag{28}$$

where s_π is a row vector that selects inflation from z_t. Similarly, mean unemployment can be approximated as

$$\bar{u}_t = s_u (I - A_{t|T})^{-1} \mu_{t|T}, \tag{29}$$

where s_u selects unemployment from z_t. Estimates of $\bar{\pi}_t$ and $\bar{u}_t$ are shown in Fig. 5. Once again, we focus on 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.

Table 3 and Figs. 6 and 7 characterize the main sources of uncertainty about these estimates. The table and figures are based on a method developed by Sims and Zha (1999)

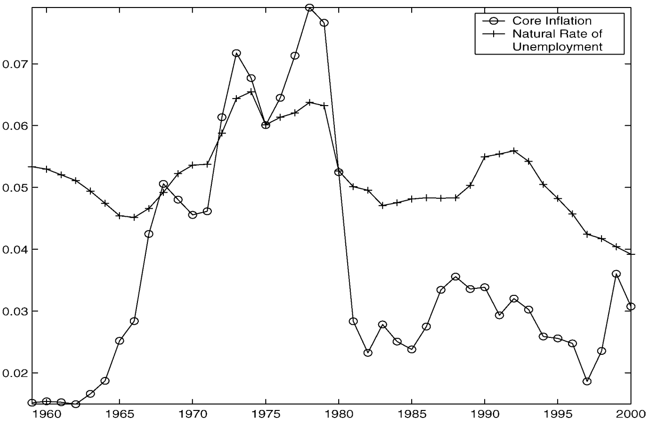

Fig. 5. Core inflation and the natural rate of unemployment.

Table 3
Principal components 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.

for constructing error bands for impulse response functions. We start by estimating the posterior covariance matrix for $\bar{\pi}_t$ via the delta method,

$$V_{\bar{\pi}} = \frac{\partial \bar{\pi}}{\partial \theta} V_\theta \frac{\partial \bar{\pi}}{\partial \theta'}. \tag{30}$$

V_θ is the $KT_A \ x \ KT_A$[13] covariance matrix for θ^T and $\partial \bar{\pi}/\partial \theta$ is the $T_A \ x \ KT_A$ matrix of partial derivatives of the function that maps VAR parameters into core inflation, evaluated at the posterior mean of θ^T. The posterior covariance V_θ was estimated from the ensemble of Metropolis draws, and derivatives were calculated numerically.[14]

Because $V_{\bar{\pi}}$ is a large object, we need a tractable way to represent the information it contains. Sims and Zha recommend error bands based on the first few principal components.[15] Let $V_{\bar{\pi}} = W \Lambda W'$, where Λ is a diagonal matrix of eigenvalues and W is an orthonormal matrix of eigenvectors. A two-sigma error band for the ith principal component is

$$\bar{\pi}_t \pm 2\lambda_i^{1/2} W_i, \tag{31}$$

where λ_i is the variance of the ith principal component and W_i is the ith column of W.

Table 3 reports the cumulative proportion of the total variation accounted for by the principal components. 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, a single factor would represent uncertainty about

[13] K is the number of elements in θ, and T_A represents the number of years. We focused on every fourth observation to keep V to a manageable size.

[14] We used this roundabout method of approximating $V_{\bar{\pi}}$ because the direct estimate was contaminated by a few outliers that 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.

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

the terminal estimate, but smoothed estimates going backward in time would be perfectly correlated with the terminal estimate and would contribute no additional uncertainty.[16] V would be a $T_A x T_A$ matrix 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 signifies that there is also substantial variation in the shape of the path.

Error bands for core inflation are shown in Fig. 6. The central dotted line is the posterior mean estimate, reproduced from Fig. 5. The horizontal line is a benchmark, end-of-sample, time-invariant estimate of mean inflation.

The first principal component accounts for roughly half the total variation and describes uncertainty about 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 is related to changes in inflation persistence. Core inflation can be interpreted as a long-horizon forecast, and

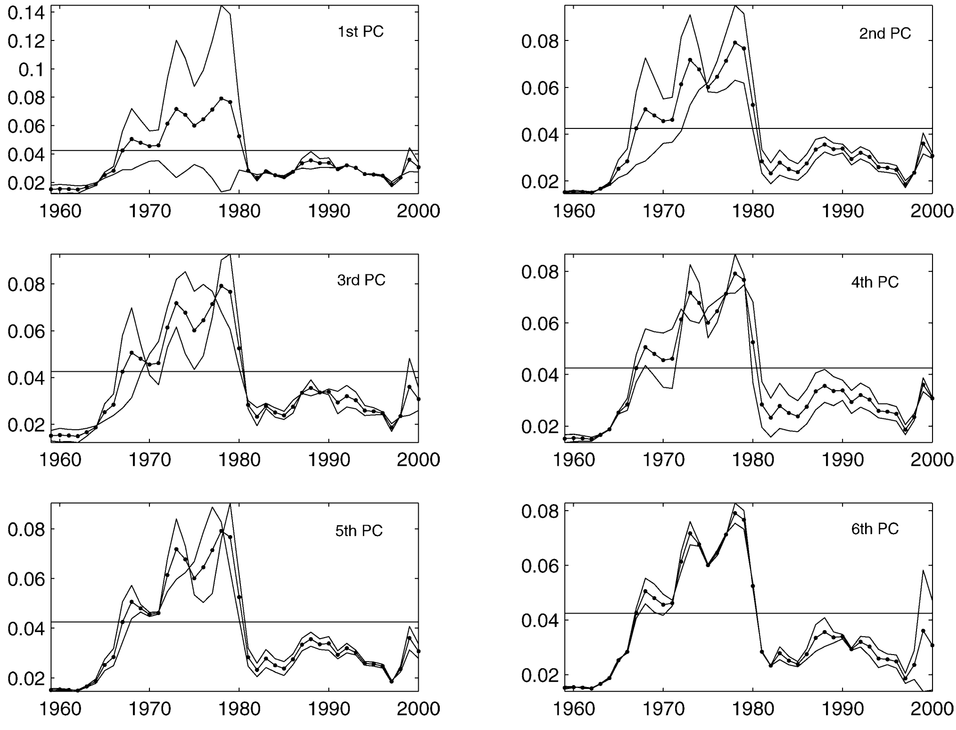

Fig. 6. Two-sigma error bands for core inflation.

[16] 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 θ_{t+1} and θ_t.

the variance of long-horizon forecasts depends positively on the degree of persistence. As shown below, inflation became more persistent as core inflation rose, and 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.

Error bands for the natural rate are constructed in the same way, and they are shown in Fig. 7. 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 $\bar{u}_t$ 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 seems to

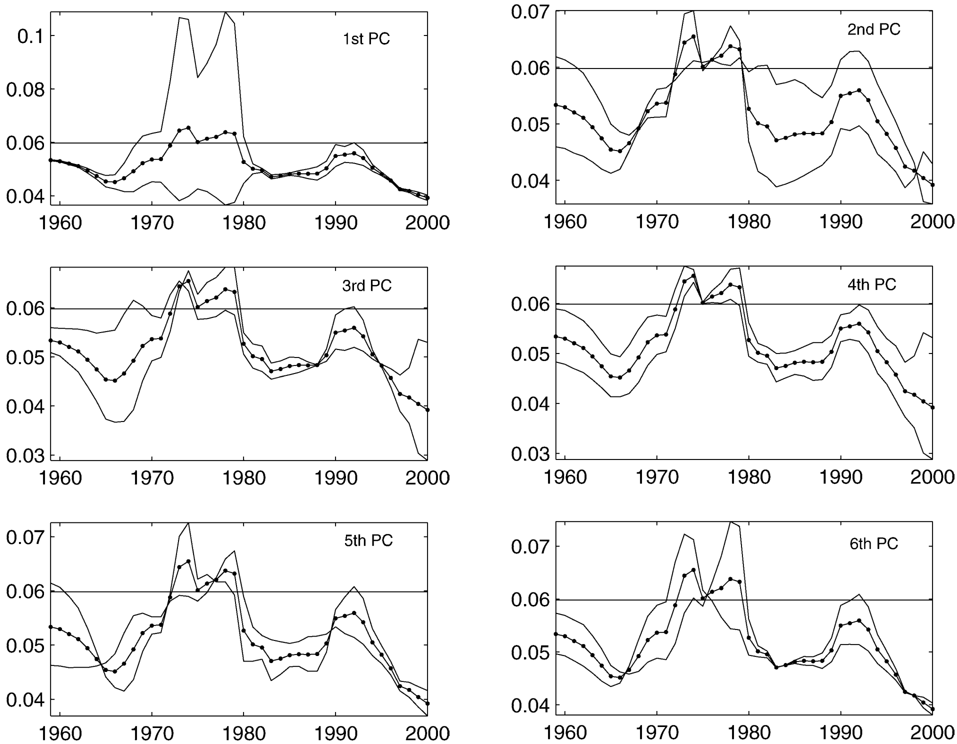

Fig. 7. Two-sigma error bands for the natural rate of unemployment.

be related to changes in the persistence of 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 the evolution of second moments of inflation, which are measured by a local linear approximation to the spectrum for inflation,

$$f_{\pi\pi}(\omega,t) = s_\pi \left(I - A_{t|T}\mathrm{e}^{-i\omega}\right)^{-1} \frac{E(R_{t|T})}{2\pi} \left(I - A_{t|T}\mathrm{e}^{i\omega}\right)^{-1\prime} s_\pi', \tag{32}$$

evaluated at the posterior mean of θ and R. An estimate of $f_{\pi\pi}(\omega,t)$ is shown in Fig. 8. Time is plotted on the x-axis, frequency on the y-axis, and power on the z-axis. Again, the estimates are similar to those reported in Cogley and Sargent (2001). The introduction of drift in R_t does not undermine our earlier evidence about 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,

$$g_{\pi\pi}(\omega,t) = \frac{f_{\pi\pi}(\omega,t)}{\int_{-\pi}^{\pi} f_{\pi\pi}(\omega,t)\,d\omega}, \tag{33}$$

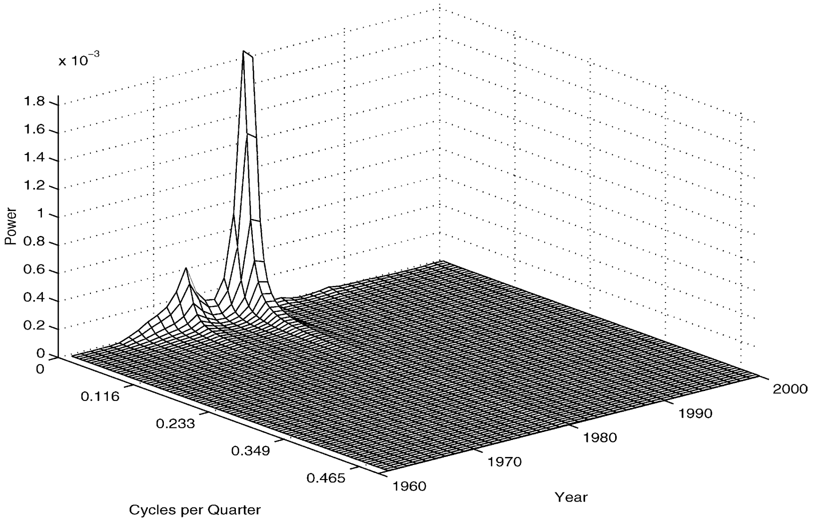

Fig. 8. Spectrum for inflation

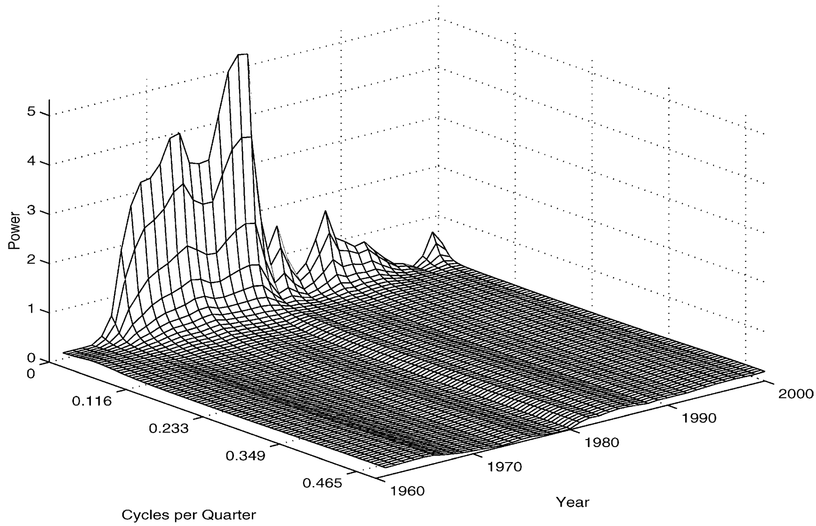

Fig. 9. Normalized spectrum for inflation.

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 Fig. 9. As in Fig. 8, 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 10 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 $V_{\bar{\pi}}$ and $V_{\bar{u}}$. 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 $g_{\pi\pi}(0,t)$.

Error bands for the first two components load heavily on the 1970s. Although the bands suggest greater persistence than in the early 1960s or mid-1990s, the exact magnitude of the increase is hard to pin down. Roughly speaking, the first two error bands 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 $g_{\pi\pi}(0,t)$. 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

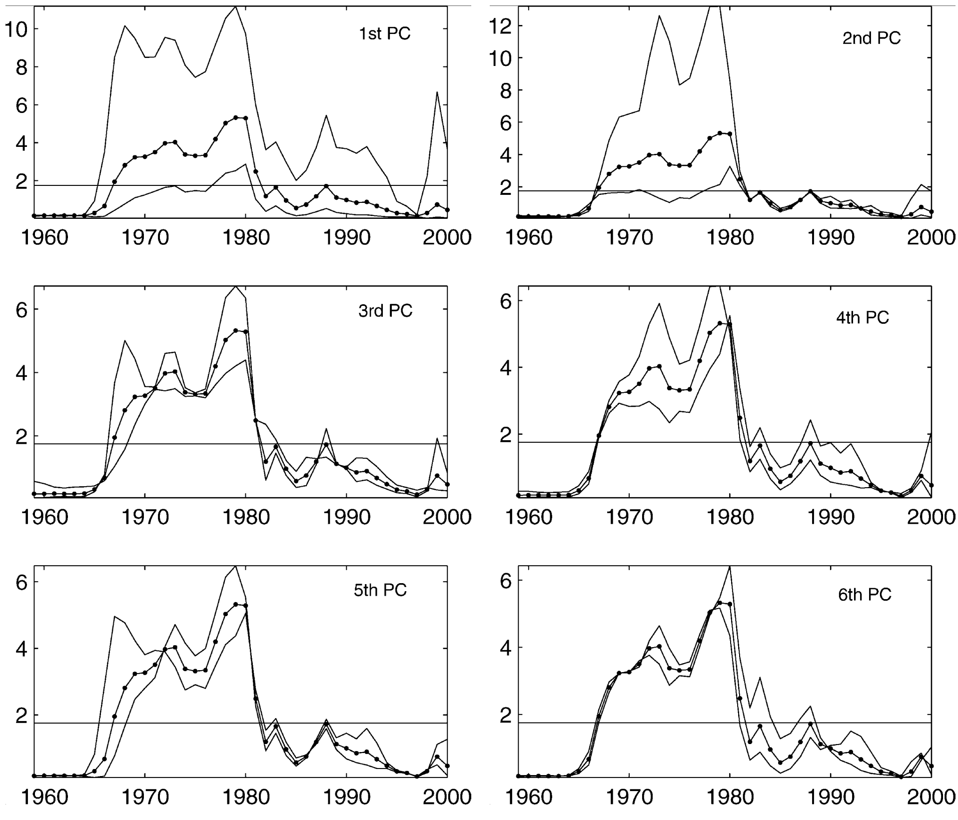

Fig. 10. Two-sigma error bands for the normalized spectrum at zero.

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.

The relation between the two series is illustrated in Fig. 11, which reproduces estimates from Figs. 4 and 8. Inflation became more persistent as core inflation rose in the 1960s and 1970s, and both features fell sharply during the Volcker disinflation. This correlation is problematic for the escape route models of Sargent (1999) and Cho et al. (2002), which predict that inflation persistence should grow along the transition from high to low inflation. Our estimates suggest the opposite pattern. Cogley and Sargent (2003) and Primiceri (2003b) formulate alternative models of central bank learning that deliver a positive correlation. They describe how during the 1970s policymakers became reluctant to disinflate rapidly.

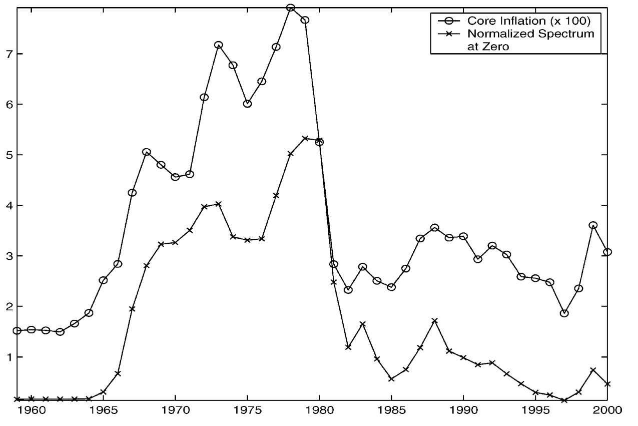

Fig. 11. Core inflation and inflation persistence.

3.6.3. Monetary policy activism

Finally, we consider evidence on the evolution of policy activism.[17] Following Clarida et al. (2000), we estimate this from a forward-looking Taylor rule with interest rate smoothing,

$$i_t = \beta_0 + \beta_1 E_t \bar{\pi}_{t,t+h_\pi} + \beta_2 E_t \bar{u}_{t,t+h_u} + \beta_3 i_{t-1} + \eta_t, \tag{34}$$

where $\bar{\pi}_{t,t+h_\pi}$ represents average inflation from t through $t + h_\pi$ and $\bar{u}_{t,t+h_u}$ is average unemployment. The activism parameter is defined as $A = \beta_1(1 - \beta_3)^{-1}$, and the policy rule is said to be activist if $A \geqslant 1$. In some models, an activist monetary rule delivers a determinate equilibrium that eradicates sunspots as determinants of 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 of average inflation and unemployment, $E_t \bar{\pi}_{t,t+h_\pi}$ and $E_t \bar{u}_{t,t+h_u}$, onto a set of instruments. The second involves projecting current interest rates onto the fitted values from the first-stage projection along with the lagged nominal interest rate. At each date, we parameterize the VAR with posterior mean estimates of θ_t and R_t and calculate population projections associated with those values.[18]

[17] This terminology follows Leeper (1991), who classifies a monetary policy rule as 'active' if it calls for a more-than-proportional increase in the nominal interest rate when inflation rises. Thus, an activist policy responds to higher inflation by increasing the real interest rate. In contrast, a policy rule is said to be 'passive' if the nominal interest rate rises less than one-for-one with inflation, so that the real interest rate is allowed to fall when inflation increases. See Woodford (2003) for a discussion of the connection between policy activism and the determinacy of equilibrium.

[18] The first-stage projections are just the average of forecasts from steps 1 through h_π and h_u, respectively, for inflation and unemployment. These are linear combinations of the right hand variables, $\phi'_{\pi t} x_t$ and $\phi'_{ut} x_t$. In the second stage, we project i_t onto those forecasts plus lagged nominal interest. To calculate the second-stage projection, we form the spectral density matrix for y_t associated with the posterior mean estimates of θ_t and R_t, then use that to deduce the spectrum for $[\phi'_{\pi t} x_t, \phi'_{ut} x_t, i_{t-1}]$ and its cross-spectrum with i_t. After integrating across frequencies, we have the variance matrix for $[\phi'_{\pi t} x_t, \phi'_{ut} x_t, i_{t-1}]$, V_{xx}, and its covariance matrix with i_t, V_{xy}. The second-stage projection is just $V_{xx}^{-1} V_{xy}$.

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 $E_t\bar{\pi}_{t,t+h_\pi}$ and $E_t\bar{u}_{t,t+h_u}$ are the VAR forecasts $E_{t-1}\bar{\pi}_{t,t+h_\pi}$ and $E_{t-1}\bar{u}_{t,t+h_u}$, respectively.

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.'s approach relative to structural VAR methods. Establishing that the Fed observes *some* variables is easier than compiling a complete list of what the Fed sees.

Our estimation strategy does impose cross-equation restrictions on the VAR, however, since it relates one-step ahead forecasts of the nominal interest rate to averages of multi-step forecasts of inflation and unemployment. We checked these cross-equation restrictions by comparing one-step ahead VAR forecasts for the interest rate with those implied by the estimated Clarida et al.'s rule, and we found that the two forecasts track one another very closely. The mean difference between the two is only 1 basis point at an annual rate, and the standard deviation is only 8 basis points. The VAR predictions are marginally better, with an innovation standard deviation of 0.877 versus 0.879 for the Clarida et al.'s rule, but the difference is in the third decimal point. Thus, the cross-equation restrictions seem appropriate.

Estimates of A are shown in Fig. 12. Here we assume $h_\pi = 4$ and $h_u = 2$, a specification motivated by Federal Reserve folk wisdom about policy lags. According to that folk wisdom, there is a lag of roughly 6 months between movements in the funds rate and the response of real variables such as GDP or unemployment, and a lag of roughly 12 months in the reaction of inflation. Much of the discussion within the Federal Reserve therefore focuses on forecasts of inflation and unemployment up to those horizons, and our bench-

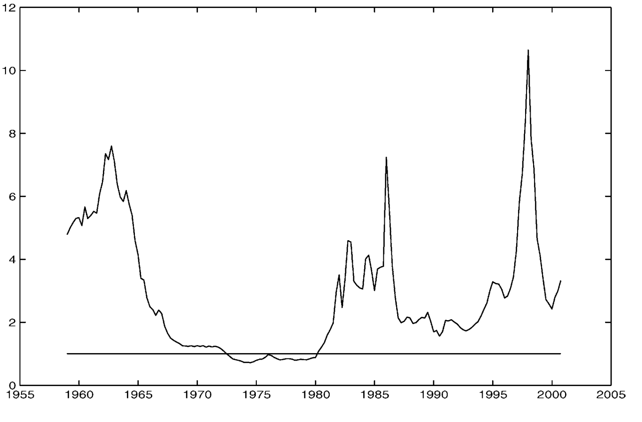

Fig. 12. Estimates of the activism coefficient.

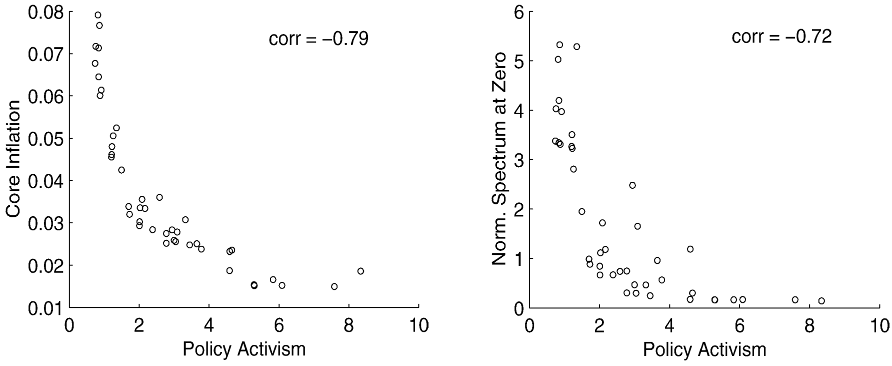

Fig. 13. Policy activism, core inflation, and inflation persistence.

mark specification is motivated by that practice. We also comment below on two other specifications, noting how the evidence differs from our benchmark case.

The estimates shown in Fig. 12 broadly resemble those reported by Clarida et al., as well as those in our earlier paper. Judging by point estimates, the policy rule was activist in the early 1960s, became approximately neutral in the late 1960s, turned passive in the early 1970s, and remained so until the early 1980s. The degree of activism rose sharply around the time of the Volcker disinflation and has remained above 1 ever since. As shown in Fig. 13, the estimates of A are inversely related to core inflation and the normalized spectrum at zero, suggesting that changes in policy activism may have contributed to the rise and fall of inflation as well as to changes in its persistence.

Figure 14 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 more than one component contributes to V_A is evidence of variation in the path of A. But the shape of the path is well determined only in the middle of the sample. The first four components record substantial uncertainty at the beginning and end. We interpret this as a symptom of weak identification.[19] Uncertainty about A is greatest 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. This makes inferences about A fragile at the beginning and end of the sample. There is better evidence of changes in A in the middle of the sample, when inflation and unemployment were more persistent. Lagged variables were more relevant as instruments at such times, making estimates of A more precise.

The next figure (Fig. 15) provides more details about the posterior for A_t in the Burns and Volcker–Greenspan terms, showing histograms for A_t for the years 1975, 1985, and 1995. Histograms were constructed by calculating an activism parameter for each draw of

[19] See Sims and Zha's discussion of identification of forward-looking Taylor rules.

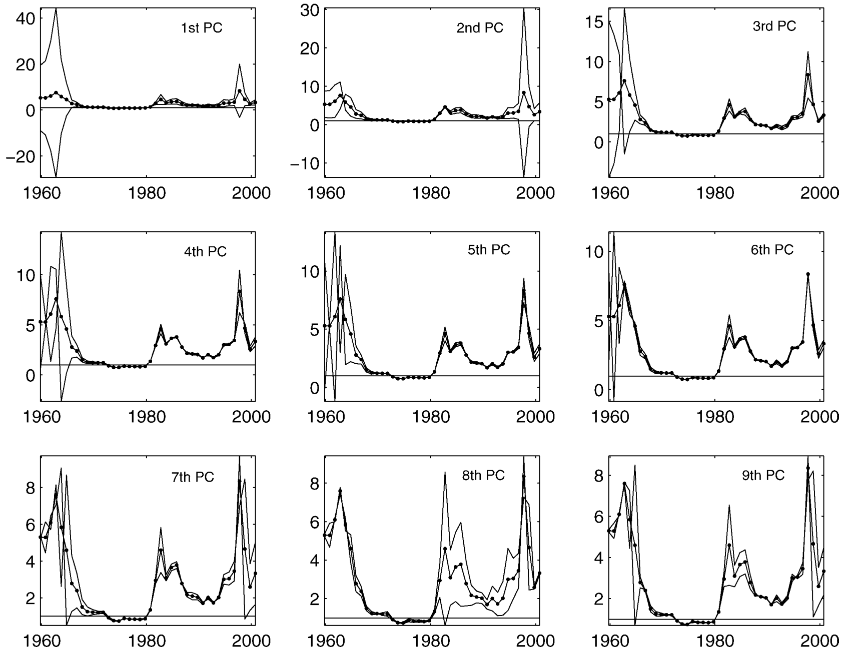

Fig. 14. Two-sigma error bands for the activism parameter.

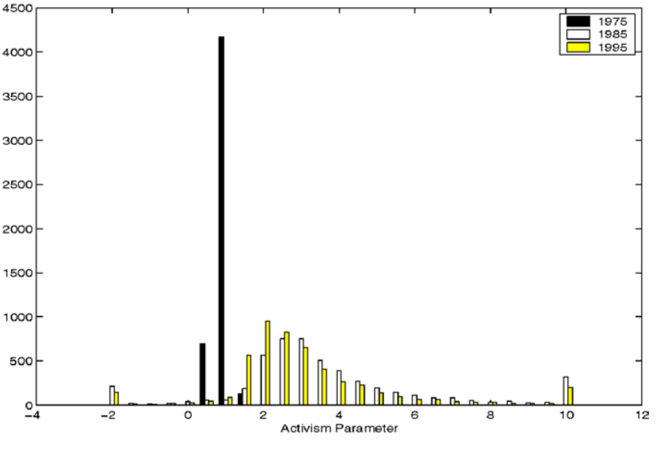

Fig. 15. Histograms for A_t in selected years.

θ_t and R_t in our simulation, for a total of 5000 in each year. Values for 1975 are shown in black, those for 1985 are in white, and estimates for 1995 are shown in gray; outliers are collected in the end bins.

In 1975, the probability mass was concentrated near 1, and there was a 0.208 probability that $A_t > 1$. By 1985, the center of the distribution had shifted to the right, and the probability that $A_t > 1$ had risen to 0.919. The distribution for 1995 is 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. We interpret this as strong—though perhaps not dispositive—evidence of greater policy activism.

We also considered two other specifications, a forward-looking rule with short forecasting horizons ($h_\pi = h_u = 1$) and a conventional Taylor rule with interest-rate smoothing. The numbers for the short-horizon forward-looking rule are similar to those for its longer-term cousin. The probability that $A_t > 1$ was 0.131 in 1975, 0.911 in 1985, and 0.896 in 1995, and A_t was higher in the 1980s and 1990s along 92 percent of the sample paths.

Results for the Taylor rule are similar in some respects, different in others. Like the forward-looking rules, the estimates suggest a passive rule in the mid-1970s and a more active stance in the mid-1980s. For example, there was only a 6.7 percent probability that $A_t > 1$ in 1975 but a 92 percent probability in 1985. Comparing points on the same trajectories, we find a 97 percent probability of an increase in A_t from the mid-1970s to the mid-1980s. One notable difference is that Taylor-rule estimates suggest the Federal Reserve was backsliding in the mid-1990s, having moved toward a policy that was approximately neutral. According to this specification, there was only a 50–50 chance that $A_t > 1$ in 1995, and activism had increased relative to 1975 on only two thirds of the sample paths. Perhaps this reflects the Fed's flirtation with an 'opportunistic' approach to disinflation at the time.

On balance, the evidence on policy activism seems to corroborate the conclusion of 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 because of instrument relevance problems, and the tails of the distributions at various dates overlap. But it seems clear that the center of the distribution for A_t has shifted to the right, from a position below 1 in the mid-1970s to a level greater than one for most of the period after the early 1980s.

3.7. *Tests for θ stability*

Finally, we consider classical tests for variation in θ. Bernanke and Mihov (1998a, 1998b) 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 θ. For reduced form vector autoregressions similar to ours, they were unable to reject the hypothesis of time invariance.

We applied their test to our data and reaffirmed their results. The first row of Table 4 summarizes our 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. Columns labeled with variable names refer to the single-equation tests, and the column labeled 'VAR' refers to the system-wide test. In each case, we cannot reject that θ is time invariant.[20]

[20] 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.

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

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

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

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 Eq. (1) with draws of θ^T, H^T, and B from the posterior density. For each draw of (θ^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 θ 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 the highest success rate, detecting drift in θ in about one-fourth of the samples. The detection probabilities are lower in the single equation tests, which reject at the 5 percent level only about 14 percent of the time. Thus, even when θ 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 θ at some unknown break date. The results of this experiment may just reflect that the test is less well suited to detect alternatives like 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.

When applied to actual data, the Nyblom–Hansen test also fails to reject time invariance for θ. 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.

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 notes to Table 4.

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.

Boivin (1999) conjectures that the sup-Wald version of Andrews's test has higher power than the others, 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 that we study in which the detection probability exceeds 50 percent. Notice, however, 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.

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 θ is indeed a manifestation of changes in monetary policy, with all other features remaining constant, then tests for stability of the policy-rule parameters should be more powerful than for stability of the VAR as a whole. The vector θ has high dimension, and the drifting components in θ should lie in a lower-dimensional subspace corresponding to drifting policy parameters, if monetary policy is the source of instability.

To examine the evidence on stability of the Clarida et al.'s rule, we replicated their GMM estimator using our data and instruments, and then calculated Andrews's statistics to test for time-invariance.[21] Notice that here we are adopting a different approach from that taken earlier in the paper. Now we are estimating time-invariant policy rules for subsamples, as in Clarida et al., and probing for evidence of parameter shifts across subsamples. But instead of selecting potential break dates a priori, we follow Andrews and explore all possible break dates in the middle 70 percent of the sample. Thus we are studying the power of Andrews's tests when applied to the GMM estimator of Clarida et al. The full sample is still 1959–2000,[22] and potential break dates are permitted between 1965 and 1994.

Perhaps surprisingly in light of the results of Clarida et al., the tests fail to reject invariance across subsamples (see Table 7). We repeated the procedure for artificial samples generated from our VAR to check the power of the tests. Once again, the tests have low detection probabilities.

We also tried to increase power by concentrating on a single linear combination of θ 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 Figs. 16 and 17 illustrate, the first principal component dominates the variation in $\theta_{t|T}$;[23] most of the other principal components are

[21] We chose Andrews's tests because the CCG rule is estimated by GMM. The Nyblom–Hansen test is based on ML estimates.

[22] Recall that data for 1948–1958 were used to set priors; the sample 1959–2000 is the same as that used above.

[23] More precisely, the figures illustrate partial sums of the first principal component for $\Delta\theta_{t|T}$.

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.

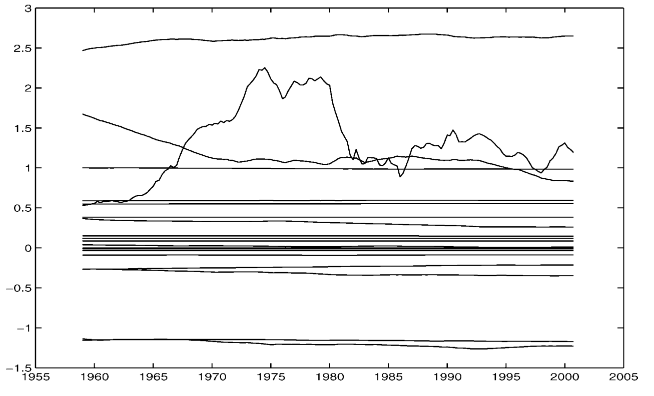

Fig. 16. Principal components of $\theta_{t|T}$.

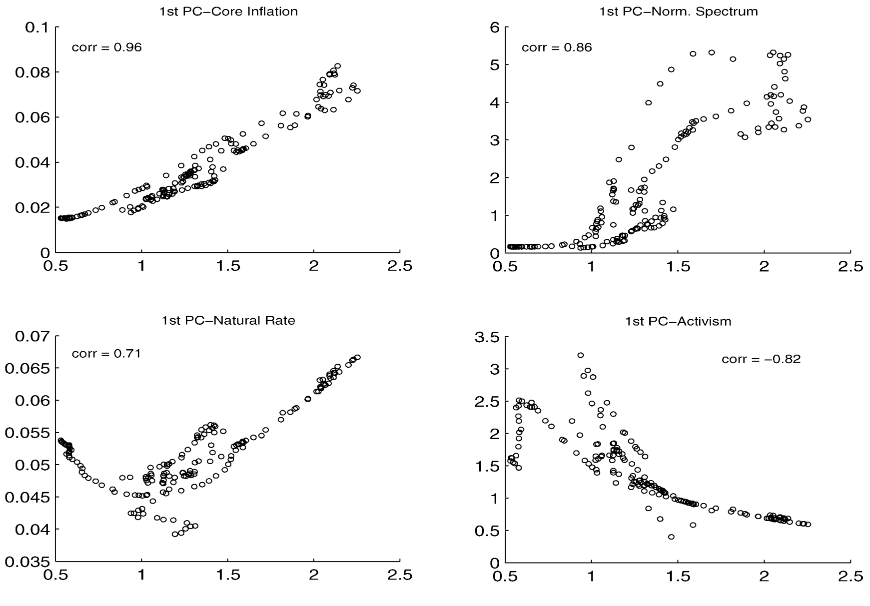

Fig. 17. Correlation of first principal component with other features.

Table 8
Stability of the first principal component

	sup-LM	sup-Wald
Data	F	R 5%
Power	0.220	0.087

See the note to Table 4.

approximately time invariant. The first component is also highly correlated with variation in the features discussed above, so it seems a promising candidate on which to concentrate.

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 θ seems better. Despite the low power, one of the tests rejects time invariance in actual data.

To summarize, most of our tests fail to reject time invariance of θ, 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 θ, for the data time invariance is rejected for the US data at the one-percent level. One reasonable interpretation is that θ is drifting, but that most of the procedures are unable to detect it. Still, the main point here is not that one test rejects time invariance, but that the majority of tests that fail to reject have low power. Accordingly, a failure to reject should not be construed as an embarrassment to time-varying parameter models.

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 θ often falls in the indeterminate range.

4. Conclusion

One respectable view is that either his erroneous model, his 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 US history; and that his improved model, more patience, or greater discipline led Paul Volcker to administer monetary policy in a way that conquered American inflation.[24] 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.

[24] See J. Bradford DeLong (1997) and John Taylor (1997).

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. For reasons discussed 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 prudent to be cautious in interpreting evidence either for or against drifting coefficients. For reasons that are clearest in continuous time (see Anderson et al., 2003), 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 that implement Bernanke and Mihov's tests under an artificial economy with drifting coefficients.

Acknowledgments

For comments and suggestions, we thank Jean Boivin, Fabio Canova, Marco Del Negro, Mark Gertler, Sergei Morozov, Adrian Pagan, Simon Potter, Christopher Sims, Mark Watson, Tao Zha, an anonymous referee, and participants at numerous seminars and conferences.

Appendix A. The relation between the restricted and unrestricted models

A.1. Priors

The prior for the unrestricted model is

$$f\big(\theta^T, Q, \Omega\big) = f\big(\theta^T \mid Q, \Omega\big) f(Q, \Omega), \tag{35}$$

where θ^T represents the VAR parameters, Q is their innovation variance, and Ω stands for everything else. Because of the independence assumptions on the prior, this can be written as

$$f\big(\theta^T, Q, \Omega\big) = f\big(\theta^T \mid Q\big) f(Q) f(\Omega). \tag{36}$$

The restricted model adds an a priori condition that rules out explosive values of θ,

$$p\big(\theta^T, Q, \Omega\big) = \frac{I(\theta^T) f(\theta^T, Q, \Omega)}{\iiint I(\theta^T) f(\theta^T, Q, \Omega)\, \mathrm{d}\theta^T\, \mathrm{d}Q\, \mathrm{d}\Omega}. \tag{37}$$

Thus, the stability condition truncates and renormalizes the unrestricted prior.

We can factor $f(\theta^T, Q, \Omega)$ as before to obtain

$$p\big(\theta^T, Q, \Omega\big) = \frac{I(\theta^T) f(\theta^T \mid Q) f(Q) f(\Omega)}{\iiint I(\theta^T) f(\theta^T \mid Q) f(Q) f(\Omega)\, \mathrm{d}\theta^T\, \mathrm{d}Q\, \mathrm{d}\Omega}$$

$$= \frac{I(\theta^T)f(\theta^T \mid Q)f(Q)f(\Omega)}{[\int f(\Omega)\,\mathrm{d}\Omega]\iint I(\theta^T)f(\theta^T \mid Q)f(Q)\,\mathrm{d}\theta^T\,\mathrm{d}Q}$$

$$= \frac{I(\theta^T)f(\theta^T \mid Q)f(Q)f(\Omega)}{\iint I(\theta^T)f(\theta^T \mid Q)f(Q)\,\mathrm{d}\theta^T\,\mathrm{d}Q}, \tag{38}$$

where the last equality follows from the fact that $f(\Omega)$ is proper. Now define

$$m_\theta(Q) \equiv \int I(\theta^T) f(\theta^T \mid Q)\,\mathrm{d}\theta^T, \tag{39}$$

and

$$m_Q \equiv \int m_\theta(Q) f(Q)\,\mathrm{d}Q. \tag{40}$$

The term $m_\theta(Q)$ is the conditional probability of a non-explosive draw from the unrestricted transition density, $f(\theta^T \mid Q)$, as a function of Q. The number m_Q is the mean of the conditional probabilities, averaged across draws from the marginal prior $f(Q)$. Since both are probabilities, it follows that

$$0 \leqslant m_\theta(Q) \leqslant 1, \qquad 0 \leqslant m_Q \leqslant 1. \tag{41}$$

The left-hand inequality for m_Q is strict if there is some chance of a non-explosive draw for some value of Q. For finite T there always is.

After rearranging terms in (38), we find

$$p(\theta^T, Q, \Omega) = \frac{I(\theta^T)f(\theta^T \mid Q)}{m_\theta(Q)} \frac{m_\theta(Q)f(Q)}{m_Q} f(\Omega). \tag{42}$$

Thus the restricted conditional prior for θ^T given Q is

$$p(\theta^T \mid Q) = \frac{I(\theta^T)f(\theta^T \mid Q)}{m_\theta(Q)} \tag{43}$$

(i.e. Eq. (7)). Similarly, the restricted marginal prior for Q is

$$p(Q) = \frac{m_\theta(Q)f(Q)}{m_Q} \tag{44}$$

(i.e. Eq. (8)). The marginal prior for Ω remains the same as for the unrestricted model, $p(\Omega) = f(\Omega)$. Notice that each term is normalized to integrate to 1; i.e., each component is proper.

From (7) we can derive the restricted transition density. This is defined as

$$p(\theta_{t+1} \mid \theta_t, Q) = \frac{p(\theta_{t+1}, \theta_t \mid Q)}{p(\theta_t \mid Q)}. \tag{45}$$

The numerator can be expressed as

$$p(\theta_{t+1}, \theta_t \mid Q) = \iint p(\theta^T \mid Q)\,\mathrm{d}\theta^{t-1}\,\mathrm{d}\theta^{t+2,T}, \tag{46}$$

where θ^{t-1} represents the history of θ_s up to date $t-1$ and $\theta^{t+2,T}$ represents the path from dates $t+2$ to T. After substituting from Eq. (7), this becomes

$$p(\theta_{t+1}, \theta_t \mid Q) = \frac{1}{m_\theta(Q)} \iint \prod_{s=0}^{T-1} I(\theta_{s+1}) f(\theta_{s+1} \mid \theta_s, Q)\,\mathrm{d}\theta^{t-1}\,\mathrm{d}\theta^{t+2,T}. \tag{47}$$

The integrand can be expanded as

$$\begin{aligned}\prod_{s=0}^{T-1} I(\theta_{s+1}) f(\theta_{s+1} \mid \theta_s, Q) = &\prod_{s=0}^{t-1} I(\theta_{s+1}) f(\theta_{s+1} \mid \theta_s, Q) \\ &\times I(\theta_{t+1}) f(\theta_{t+1} \mid \theta_t, Q) \\ &\times \prod_{s=t+1}^{T-1} I(\theta_{s+1}) f(\theta_{s+1} \mid \theta_s, Q). \end{aligned} \tag{48}$$

It follows that

$$\begin{aligned} p(\theta_{t+1}, \theta_t \mid Q) = &\frac{1}{m_\theta(Q)} \int \prod_{s=0}^{t-1} I(\theta_{s+1}) f(\theta_{s+1} \mid \theta_s, Q)\, \mathrm{d}\theta^{t-1} \\ &\times I(\theta_{t+1}) f(\theta_{t+1} \mid \theta_t, Q) \\ &\times \int \prod_{s=t+1}^{T-1} I(\theta_{s+1}) f(\theta_{s+1} \mid \theta_s, Q)\, \mathrm{d}\theta^{t+2,T}. \end{aligned} \tag{49}$$

The marginal density for θ_t can be expressed as

$$\begin{aligned} p(\theta_t \mid Q) &= \int p(\theta_{t+1}, \theta_t \mid Q)\, \mathrm{d}\theta_{t+1} \\ &= \frac{1}{m_\theta(Q)} \int \prod_{s=0}^{t-1} I(\theta_{s+1}) f(\theta_{s+1} \mid \theta_s, Q)\, \mathrm{d}\theta^{t-1} \\ &\quad \times \int \prod_{s=t}^{T-1} I(\theta_{s+1}) f(\theta_{s+1} \mid \theta_s, Q)\, \mathrm{d}\theta^{t+1,T}. \end{aligned} \tag{50}$$

The transition density is the ratio of (49) to (49):

$$\begin{aligned} &p(\theta_{t+1} \mid \theta_t, Q) \\ &\quad = \frac{I(\theta_{t+1}) f(\theta_{t+1} \mid \theta_t, Q) \int \prod_{s=t+1}^{T-1} I(\theta_{s+1}) f(\theta_{s+1} \mid \theta_s, Q)\, \mathrm{d}\theta^{t+2,T}}{\int \prod_{s=t}^{T-1} I(\theta_{s+1}) f(\theta_{s+1} \mid \theta_s, Q)\, \mathrm{d}\theta^{t+1,T}} \\ &\quad = \frac{I(\theta_{t+1}) f(\theta_{t+1} \mid \theta_t, Q) \int I(\theta^{t+2,T}) f(\theta^{t+2,T} \mid \theta_{t+1}, Q)\, \mathrm{d}\theta^{t+2,T}}{\int I(\theta_{t+1}) f(\theta_{t+1} \mid \theta_t, Q) [\int I(\theta^{t+2,T}) f(\theta^{t+2,T} \mid \theta_{t+1}, Q)\, \mathrm{d}\theta^{t+2,T}]\, \mathrm{d}\theta_{t+1}}. \end{aligned} \tag{51}$$

The integral in the numerator is the expectation of $I(\theta^{t+2,T})$ with respect to the conditional density $f(\theta^{t+2,T} \mid \theta_{t+1}, Q)$. This represents the probability that random walk trajectories emanating from θ_{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

$$p(\theta_{t+1} \mid \theta_t, Q) = \frac{I(\theta_{t+1}) f(\theta_{t+1} \mid \theta_t, Q) \pi(\theta_{t+1}, Q)}{\int I(\theta_{t+1}) f(\theta_{t+1} \mid \theta_t, Q) \pi(\theta_{t+1}, Q)\, \mathrm{d}\theta_{t+1}}. \tag{52}$$

A.2. Posteriors

The posterior for the restricted model is

$$p\big(\theta^T, Q, \Omega \mid Y^T\big) = \frac{p(Y^T \mid \theta^T, Q, \Omega)p(\theta^T \mid Q)p(Q)p(\Omega)}{m(Y^T)}, \tag{53}$$

where $m(Y^T)$ is the marginal likelihood. After substituting from Eqs. (7) and (8), we can express this as

$$\begin{aligned} p\big(\theta^T, Q, \Omega \mid Y^T\big) &= \frac{p(Y^T \mid \theta^T, Q, \Omega)}{m(Y^T)} \frac{I(\theta^T) f(\theta^T \mid Q)}{m_\theta(Q)} \frac{m_\theta(Q) f(Q)}{m_Q} f(\Omega), \\ &= \frac{I(\theta^T)[p(Y^T \mid \theta^T, Q, \Omega) f(\theta^T \mid Q) f(Q) f(\Omega)]}{m(Y^T) m_Q}. \end{aligned} \tag{54}$$

The term in brackets in the numerator is the posterior kernel for the unrestricted model. After multiplying and dividing by the marginal likelihood for the unrestricted model, which we denote $m_U(Y^T)$, we find

$$p\big(\theta^T, Q, \Omega \mid Y^T\big) = \frac{m_U(Y^T)}{m(Y^T) m_Q} I\big(\theta^T\big) p_U\big(\theta^T, Q, \Omega \mid Y^T\big), \tag{55}$$

where $p_U(\theta^T, Q, \Omega \mid Y^T)$ is the posterior corresponding to the unrestricted prior, $f(\cdot)$. The posterior for the restricted model is proportional to the truncation of the posterior of the unrestricted model, with a factor of proportionality depending on the normalizing constants $m_Q, m_L(Y^T)$, and $m(Y^T)$.

Appendix B. A Markov Chain Monte Carlo algorithm for simulating the posterior density

We use MCMC methods to simulate the restricted posterior density. As in our earlier paper, we simulate the unrestricted posterior $p_U(\cdot \mid Y^T)$, and then use rejection sampling to rule out explosive outcomes. The first part of this appendix justifies rejection sampling, and the second describes the algorithm used for simulating draws from $p_U(\cdot \mid Y^T)$.

B.1. Rejection sampling

The target density is $p(\theta^T, Q, \Omega \mid Y^T)$, and the proposal is $p_U(\theta^T, Q, \Omega \mid Y^T)$. Since the former is proportional to a truncation of the latter, the proposal is well-defined and positive on the support of the target. Since the proposal is a probability density, it integrates to 1. The importance ratio,

$$\begin{aligned} R\big(\theta^T, Q, \Omega\big) &= \frac{p(\theta^T, Q, \Omega \mid Y^T)}{p_U(\theta^T, Q, \Omega | Y^T)} \\ &= \frac{m_U(Y^T)}{m_Q m(Y^T)} I\big(\theta^T\big) \end{aligned}$$

$$\leqslant \frac{m_U(Y^T)}{m_Q m(Y^T)} \equiv \bar{R}, \tag{56}$$

has a known, finite upper bound, $\bar{R}$. The acceptance probability is

$$\frac{R(\theta^T, Q, \Omega)}{\bar{R}} = I(\theta^T). \tag{57}$$

This says we accept if θ^T is non-explosive and reject otherwise. Thus, we can sample from the posterior of the restricted model by simulating the unrestricted model and discarding the explosive draws.

B.2. *Sampling from* $p_U(\cdot \mid Y^T)$

We combine the techniques used in Cogley and Sargent (2001) with those of Jacquier et al. (1994) to construct a Metropolis-within-Gibbs sampler. The algorithm consists of 5 steps, one for θ^T, Q, β, the elements of σ, 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, β is treated as a vector of regression parameters, and the elements of σ 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.2.1. *VAR parameters,* θ^T

We first consider the distribution of VAR parameters conditional on the data and other blocks of parameters. Conditional on H^T and β, 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[25]

$$p_U(\theta^T \mid Y^T, Q, R^T) = f(\theta_T \mid Y^T, Q, R^T) \prod_{t=1}^{T-1} f(\theta_t \mid \theta_{t+1}, Y^t, Q, R_t). \tag{58}$$

The unrestricted model is a linear, conditionally Gaussian state-space model. Assuming a Gaussian prior for θ_0, all the conditional densities on the right-hand side of (58) are Gaussian. Their means and variances can be computed via a forward and backward recursion.

The forward recursion uses the Kalman filter. Let

$$\begin{aligned} \theta_{t|t} &\equiv E(\theta_t \mid Y^t, Q, R^T), \\ P_{t|t-1} &\equiv \mathrm{Var}(\theta_t \mid Y^{t-1}, Q, R^T), \\ P_{t|t} &\equiv \mathrm{Var}(\theta_t \mid Y^t, Q, R^T), \end{aligned} \tag{59}$$

represent conditional means and variances going forward in time. These can be computed recursively, starting from the prior mean and variance for θ_0:

[25] The elements of σ are redundant conditional on H^T.

$$\begin{aligned}
K_t &= P_{t|t-1} X_t \big(X_t' P_{t|t-1} X_t + R_t\big)^{-1}, \\
\theta_{t|t} &= \theta_{t-1|t-1} + K_t \big(y_t - X_t' \theta_{t-1|t-1}\big), \\
P_{t|t-1} &= P_{t-1|t-1} + Q, \\
P_{t|t} &= P_{t|t-1} - K_t X_t' P_{t|t-1}.
\end{aligned} \tag{60}$$

At the end of the sample, the forward recursion delivers the mean and variance for θ_T, and this pins down the first term in (58),

$$f\big(\theta_T \mid Y^T, Q, R^T\big) = N(\theta_{T|T}, P_{T|T}). \tag{61}$$

The remaining terms in (58) are derived from a backward recursion, which updates conditional means and variances to reflect the additional information about θ_t contained in θ_{t+1}. Let

$$\begin{aligned}
\theta_{t|t+1} &\equiv E\big(\theta_t \mid \theta_{t+1}, Y^t, Q, R^T\big), \\
P_{t|t+1} &\equiv \operatorname{Var}\big(\theta_t \mid \theta_{t+1}, Y^t, Q, R^T\big),
\end{aligned} \tag{62}$$

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

$$\begin{aligned}
\theta_{t|t+1} &= \theta_{t|t} + P_{t|t} P_{t+1|t}^{-1} (\theta_{t+1} - \theta_{t|t}), \\
P_{t|t+1} &= P_{t|t} - P_{t|t} P_{t+1|t}^{-1} P_{t|t}.
\end{aligned} \tag{63}$$

The updated estimates determine the mean and variance for remaining elements in (58),

$$f\big(\theta_t \mid \theta_{t+1}, Y^T, Q, R^T\big) = N(\theta_{t|t+1}, P_{t|t+1}). \tag{64}$$

A random trajectory for θ^T is generated by iterating backward. The backward recursion starts with a draw of θ_T from (61). Then, conditional on its realization, θ_{T-1} is drawn from (64), θ_{T-2} is drawn conditional on the realization of θ_{T-1}, and so on back to the beginning of the sample.

B.2.2. 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 θ^T, the VAR parameter innovations, v_t, are observable. Furthermore, the other conditioning variables are irrelevant at this stage,

$$f\big(Q \mid Y^T, \theta^T, \sigma, \beta, H^T\big) = f\big(Q \mid Y^T, \theta^T\big). \tag{65}$$

Knowledge of σ is redundant conditional on H^T, and β and H^T are irrelevant because v_t is independent of ξ_t and η_{it}.

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

$$f\big(Q \mid Y^T, \theta^T\big) = IW\big(Q_1^{-1}, T_1\big), \tag{66}$$

with scale and degree-of-freedom parameters,

$$Q_1 = \overline{Q} + \sum_{t=1}^{T} v_t v_t' \qquad T_1 = T_0 + T. \tag{67}$$

B.2.3. Standard deviation of volatility innovations, σ

The third step involves the full conditional distribution for σ,

$$f(\sigma \mid \theta^T, H^T, \beta, Q, Y^T). \tag{68}$$

Knowledge of Q is redundant conditional on θ^T. The latter conveys information about υ_t and ε_t, but both are conditionally independent of the volatility innovations.[26] Thus, conditioning on θ^T is also irrelevant. β 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 σ_1 simplifies to

$$f(\sigma_1 \mid \sigma_2, \sigma_3, \theta^T, H^T, \beta, Q, Y^T) = f(\sigma_1 \mid h_1^T, Y^T),$$

and similarly for σ_2^2 and σ_3^2.

The scaled volatility innovations are i.i.d. normal with mean zero and variance σ_i^2. Assuming an inverse-gamma prior with scale parameter δ_0 and υ_0 degrees of freedom, the posterior is also inverse gamma,

$$f(\sigma_i^2 \mid h_i^T, Y^T) = IG\left(\frac{\upsilon_1}{2}, \frac{\delta_1}{2}\right), \tag{69}$$

where

$$\upsilon_1 = \upsilon_0 + T, \tag{70}$$

$$\delta_1 = \delta_0 + \sum_{t=1}^{T} (\Delta \ln h_{it})^2. \tag{71}$$

B.2.4. Covariance parameters, β

Next, we consider the distribution of β conditional on the data and other parameters. Knowledge of θ^T and Y^T implies knowledge of ε_t, which satisfies

$$B\varepsilon_t = u_t, \tag{72}$$

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

$$\varepsilon_{1t} = u_{1t}. \tag{73}$$

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

$$\begin{aligned} (h_{2t}^{-1/2}\varepsilon_{2t}) &= \beta_{21}(-h_{2t}^{-1/2}\varepsilon_{1t}) + (h_{2t}^{-1/2}u_{2t}), \\ (h_{3t}^{-1/2}\varepsilon_{3t}) &= \beta_{31}(-h_{3t}^{-1/2}\varepsilon_{1t}) + \beta_{32}(-h_{3t}^{-1/2}\varepsilon_{2t}) + (h_{3t}^{-1/2}u_{3t}), \end{aligned} \tag{74}$$

with independent standard normal residuals.

[26] The measurement innovations are informative for R_t, which depends indirectly on σ, but this information is subsumed in H_t.

Once again, many of the conditioning variables drop out. Q and σ are redundant conditional on θ^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,

$$f(\beta_i) = N(\beta_{i0}, V_{i0}), \quad i = 2, 3, \tag{75}$$

the posterior is also normal,

$$f\big(\beta_i \mid Y^T, \theta^T, h_i^T\big) = N(\beta_{i1}, V_{i1}), \quad i = 2, 3, \tag{76}$$

where

$$V_{i1} = \big(V_{i0}^{-1} + Z_i' Z_i\big)^{-1}, \tag{77}$$

$$\beta_{i1} = V_{i1}\big(V_{i0}^{-1}\beta_{i0} + Z_i' z_i\big). \tag{78}$$

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

B.2.5. 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, θ^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

$$f\big(h_{it} \mid h_{-it}, u_i^T, \sigma_i\big) = f\big(h_{it} \mid h_{it-1}, h_{it+1}, u_i^T, \sigma_i\big), \tag{79}$$

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 θ^T, and h_j^T and σ_j, $i \neq j$, are irrelevant because the stochastic volatilities are independent. By Bayes's theorem, the conditional kernel can be expressed as[27]

$$\begin{aligned} f\big(h_{it} \mid h_{it-1}, h_{it+1}, u_i^T, \sigma_i\big) &\propto f(u_{it} \mid h_{it}) f(h_{it} \mid h_{it-1}) f(h_{it+1} \mid h_{it}), \\ &\propto h_{it}^{-1.5} \exp\left(-\frac{u_{it}^2}{2h_{it}}\right) \exp\left(\frac{-(\ln h_{it} - \mu_{it})^2}{2\sigma_c^2}\right). \end{aligned} \tag{80}$$

Its form follows from the normal form of the conditional likelihood, $f(u_{it} \mid h_{it})$, and the log-normal form of the log-volatility equation (15). The parameters μ_{it} and σ_{ic}^2 are the conditional mean and variance of h_{it} implied by (15) and knowledge of h_{it-1} and h_{it+1}. In the random walk case, they are

$$\mu_{it} = (1/2)(\ln h_{it+1} + \ln h_{it-1}), \qquad \sigma_{ic}^2 = (1/2)\sigma_i^2. \tag{81}$$

Notice that the normalizing constant is absent from (80). Jacquier, et. al. say the normalizing constant is costly to compute, and they recommend a Metropolis step instead of a

[27] The formulas are a bit different at the beginning and end of the sample.

Gibbs step. One natural way to proceed is to draw a trial value for h_{it} from the log-normal density implied by (15), and then use the conditional likelihood $f(u_{it} \mid h_{it})$ to compute the acceptance probability. Thus, our proposal density is

$$q(h_{it}) \propto h_{it}^{-1} \exp\left(\frac{-(\ln h_{it} - \mu_{it})^2}{2\sigma_{ic}^2}\right), \tag{82}$$

and the acceptance probability for the mth draw is

$$\begin{aligned} \alpha_m &= \frac{f(u_{it} \mid h_{it}^m) q(h_{it}^m)}{q(h_{it}^m)} \frac{q(h_{it}^{m-1})}{f(u_{it} \mid h_{it}^{m-1}) q(h_{it}^{m-1})} \\ &= \frac{(h_{it}^m)^{-1/2} \exp(-u_{it}^2 / 2h_{it}^m)}{(h_{it}^{m-1})^{-1/2} \exp(-u_{it}^2 / 2h_{it}^{m-1})}. \end{aligned} \tag{83}$$

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.

References

Aguilar, O., West, M., 2000. Bayesian dynamic factor models and portfolio allocation. Journal of Business and Economic Statistics 18, 338–357.

Anderson, E., Hansen, L.P., Sargent, T.J., 2003. A quartet of semigroups for model specification, robustness, prices of risk, and model detection. Journal of the European Economic Association 1, 68–123.

Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.

Benati, L., 2001. Investigating inflation dynamics across monetary regimes: taking the Lucas Critique seriously. Unpublished manuscript. Bank of England.

Bernanke, B., Mihov, I., 1998a. The liquidity effect and long-run neutrality. Carnegie-Rochester Conference Series on Public Policy 49, 149–194.

Bernanke, B., Mihov, I., 1998b. Measuring monetary policy. Quarterly Journal of Economics 113, 869–902.

Boivin, J., 1999. Revisiting the evidence on the stability of monetary VARs. Unpublished manuscript. Graduate School of Business, Columbia University.

Cho, I.K., Williams, N., 2002. Sargent, T.J., Escaping Nash inflation. Review of Economic Studies 69, 1–40.

Clarida, R., Gali, J., Gertler, M., 2000. Monetary policy rules and macroeconomic stability: evidence and some theory. Quarterly Journal of Economics 115, 147–180.

Cogley, T., Sargent, T.J., 2001. Evolving post World War II US inflation dynamics. NBER Macroeconomics Annual 16, 331–373.

Cogley, T., Sargent, T.J., 2003. The conquest of US inflation: learning and robustness to model uncertainty. Unpublished manuscript. New York University and University of California, Davis.

DeLong, J.B., 1997. America's only peacetime inflation: the 1970s. In: Romer, C., Romer, D. (Eds.), Reducing Inflation. In: NBER Studies in Business Cycles, vol. 30.

Hansen, B.E., 1992. Testing for parameter instability in linear models. Journal of Policy Modeling 14, 517–533.

Ireland, P., 1999. Does the time-consistency problem explain the behavior of inflation in the United States? Journal of Monetary Economics 44, 279–292.

Jacquier, E., Polson, N.G., Rossi, P., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–418.

Jacquier, E., Polson, N.G., Rossi, P., 1999. Stochastic volatility: univariate and multivariate extensions. Unpublished manuscript. Finance Department, Boston College/Graduate School of Business, University of Chicago.

Kim, C.J., Nelson, C.R., 1999. Has the US economy become more stable? A Bayesian approach based on a Markov switching model of the business cycle. Review of Economics and Statistics 81, 608–661.

Leeper, E., 1991. Equilibria under 'active' and 'passive' monetary and fiscal policies. Journal of Monetary Economics 27, 129–147.
Leeper, E., Zha, T., 2001a. Empirical analysis of policy interventions. Unpublished manuscript. Department of Economics, Indiana University/Research Department, Federal Reserve Bank of Atlanta.
Leeper, E., Zha, T., 2001b. Toward a theory of modest policy interventions. Unpublished manuscript. Department of Economics, Indiana University/Research Department, Federal Reserve Bank of Atlanta.
Lindley, D., 1957. A statistical paradox. Biometrika 44, 187–192.
Lucas Jr., R.E., 1976. Econometric policy evaluation: A critique. In: Brunner, K., Meltzer, A. (Eds.), The Phillips Curve and Labor Markets. North-Holland, Amsterdam.
McCallum, B.T., 1999. Issues in the design of monetary policy rules. In: Taylor, J.B., Woodford, M. (Eds.), In: Handbook of Macroeconomics, vol. 1C. Elsevier, Amsterdam.
McConnell, M., Perez Quiros, G., 2000. Output fluctuations in the United States: What has changed since the early 1980s? American Economic Review 90, 1464–1476.
Nyblom, J., 1989. Testing for the constancy of parameters over time. Journal of the American Statistical Association 84, 223–230.
Parkin, M., 1993. Inflation in North America. In: Shigehara, K. (Ed.), Price Stabilization in the 1990s. McMillan, London.
Pitt, M., Shepard, N., 1999. Time-varying covariances: a factor stochastic volatility approach. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, vol. 6. Oxford Univ. Press, Oxford.
Primiceri, G., 2003a. Time varying structural vector autoregressions and monetary policy. Unpublished manuscript. Princeton University.
Primiceri, G., 2003b. Why inflation rose and fell: policymakers' beliefs and US postwar stabilization policy. Unpublished manuscript. Princeton University.
Romer, C.D., Romer, D.H., 2002. The evolution of economic understanding and postwar stabilization policy. In: Rethinking Stabilization Policy: A symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming August 29–31, 2002. Kansas City, pp. 11–78.
Rudebusch, G.D., Svensson, L.E.O., 1999. Policy rules for inflation targeting. In: Taylor, J.B. (Ed.), Monetary Policy Rules. Univ. of Chicago Press, Chicago. NBER conference report.
Samuelson, P.A., Solow, R.M., 1960. Analytical aspects of anti-inflation policy. American Economic Review 50, 177–184.
Sargent, T.J., 1973. The fundamental determinants of the interest rate: a comment. Review of Economics and Statistics 55, 391–393.
Sargent, T.J., 1999. The Conquest of American Inflation. Princeton Univ. Press, Princeton.
Sargent, T.J., 2002. Commentary: the evolution of economic understanding and postwar stabilization policy. In: Rethinking Stabilization Policy: A symposium sponsored by the Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming August 29–31, 2002. Kansas City, pp. 79–89.
Sims, C.A., 1980. Comparison of interwar and postwar business cycles: monetarism reconsidered. American Economic Review, 250–257.
Sims, C.A., 1999. Drifts and breaks in monetary policy. Unpublished manuscript. Department of Economics, Princeton University.
Sims, C.A., 2001. Comment on Sargent and Cogley's 'Evolving post World War II US inflation dynamics'. NBER Macroeconomics Annual 16, 373–379.
Sims, C.A., Zha, T., 1999. Error bands for impulse responses. Econometrica 67, 1113–1155.
Sims, C.A., Zha, T., 2004. Were there regime switches in US monetary policy? Unpublished manuscript. Department of Economics, Princeton University/Research Department, Federal Reserve Bank of Atlanta.
Stock, J.H., 2001. Discussion of Cogley and Sargent 'Evolving post World War II US inflation dynamics'. NBER Macroeconomics Annual 16, 379–387.
Taylor, J.B., 1993. Discretion versus policy rules in practice. Carnegie-Rochester Conference Series on Public Policy 39, 195–214.
Taylor, J.B., 1997. Comment on America's only peacetime inflation: the 1970s. In: Romer, C.D., Romer, D.H. (Eds.), Reducing Inflation, vol. 30. NBER Studies in Business Cycles.
Woodford, M., 2003. Interest and Prices. Princeton Univ. Press, Princeton.
Whittle, P., 1953. The analysis of multiple stationary time series. Journal of the Royal Statistical Society, Series B 15, 125–139.