Review of Economic Studies (2005) **72**, 821–852 0034-6527/05/00330821$02.00

Time Varying Structural Vector Autoregressions and Monetary Policy

GIORGIO E. PRIMICERI
Northwestern University and NBER

First version received December 2002; *final version accepted July* 2004 (*Eds.*)

Monetary policy and the private sector behaviour of the U.S. economy are modelled as a time varying structural vector autoregression, where the sources of time variation are both the coefficients and the variance covariance matrix of the innovations. The paper develops a new, simple modelling strategy for the law of motion of the variance covariance matrix and proposes an efficient Markov chain Monte Carlo algorithm for the model likelihood/posterior numerical evaluation. The main empirical conclusions are: (1) both systematic and non-systematic monetary policy have changed during the last 40 years—in particular, systematic responses of the interest rate to inflation and unemployment exhibit a trend toward a more aggressive behaviour, despite remarkable oscillations; (2) this has had a negligible effect on the rest of the economy. The role played by exogenous non-policy shocks seems more important than interest rate policy in explaining the high inflation and unemployment episodes in recent U.S. economic history.

1. INTRODUCTION

There is strong evidence that U.S. unemployment and inflation were higher and more volatile in the period between 1965 and 1980 than in the last 20 years. The literature has considered two main classes of explanations for this difference in performance. The first class of explanations (see, for instance, Blanchard and Simon (2001), Stock and Watson (2002), Sims and Zha (2004)) focuses on the heteroscedasticity of the exogenous shocks, which have been much more volatile in the 1970's and early 1980's than in the rest of the sample. The second class of explanations emphasizes the changes in the transmission mechanism, *i.e.* the way macroeconomic variables respond to shocks. Particular attention has been given to monetary policy. If monetary policy varies over time, this has a potential direct effect on the propagation mechanism of the innovations. Furthermore, if agents are rational and forward looking, policy changes will be incorporated in the private sector's forecasts, inducing additional modifications in the transmission mechanism.

Many authors (among others Judd and Rudebusch (1998), Clarida, Galí and Gertler (2000), Cogley and Sargent (2001, 2003), Boivin and Giannoni (2003), Lubik and Schorfheide (2004)) have argued that U.S. monetary policy was less active against inflationary pressures under the Fed chairmanship of Arthur Burns than under Paul Volcker and Alan Greenspan. However, this view is controversial. Other studies have in fact found either little evidence of changes in the systematic part of monetary policy (Bernanke and Mihov (1998), Leeper and Zha (2002), Hanson (2003)) or no evidence of unidirectional drifts in policy toward a more active behaviour (Sims, 1999, 2001*a*).

This paper investigates the potential causes of the poor economic performance of the 1970's and early 1980's and to what extent monetary policy played an important role in these high unemployment and inflation episodes. The objective here is to provide a flexible

framework for the estimation and interpretation of time variation in the systematic and non-systematic part of monetary policy and their effect on the rest of the economy. Two are the main characteristics required for an econometric framework able to address the issue: (1) time varying parameters in order to measure policy changes and implied shifts in the private sector behaviour; (2) a multiple equation model of the economy in order to understand how changes in policy have affected the rest of the economy. For this purpose, this paper estimates a time varying structural vector autoregression (VAR), where the time variation derives both from the coefficients and the variance covariance matrix of the model's innovations. Notice that any reasonable attempt to model changes in policy, structure and their interaction must include time variation of the variance covariance matrix of the innovations. This reflects both time variation of the simultaneous relations among the variables of the model and heteroscedasticity of the innovations. This is done by developing a simple multivariate stochastic volatility modelling strategy for the law of motion of the variance covariance matrix. The estimation of this model with drifting coefficients and multivariate stochastic volatility requires numerical methods. An efficient Markov chain Monte Carlo algorithm is proposed for the numerical evaluation of the posterior of the parameters of interest.

The methodology developed in the paper is used to estimate a small model of the U.S. economy, delivering many empirical conclusions. First of all, there is evidence of changes both in non-systematic and systematic monetary policy during the last 40 years. The relative importance of non-systematic policy was significantly higher in the first part of the sample, suggesting that a Taylor-type rule is much less representative of the U.S. monetary policy in the 1960's and 1970's than in the last 15 years. Furthermore, private sector responses to non-systematic policy (monetary policy shocks) appear linear in the amplitude of non-systematic policy actions. Turning to the systematic part of policy, there is some evidence of higher interest rate responses to inflation and unemployment in the Greenspan period. However, a counterfactual simulation exercise suggests these changes did not play an important role in the high inflation and unemployment episodes in recent U.S. economic history. In fact, the high volatility of the exogenous non-policy shocks seems to explain a larger fraction of the outbursts of inflation and unemployment of the 1970's and early 1980's.

From the methodological perspective, this paper is related to the fairly well developed literature on modelling and estimating time variation in multivariate linear structures. Canova (1993), Sims (1993), Stock and Watson (1996) and Cogley and Sargent (2001) model and estimate VARs with drifting coefficients. On the other hand, multivariate stochastic volatility models are discussed by Harvey, Ruiz and Shephard (1994), Jacquier, Polson and Rossi (1995), Kim, Shephard and Chib (1998), and Chib, Nardari and Shephard (2002). However, these studies impose some restrictions on the evolution over time of the elements of the variance covariance matrix. Typical restrictions are either the assumption that the covariances do not evolve independently of the variances or a factor structure for the covariance matrix. Following this line of research, Cogley (2003) and Cogley and Sargent (2003) use time varying variances in the context of VARs with drifting coefficients. However, in their model the simultaneous relations among variables are time invariant. As will be made clear in the next section, their analysis is limited to reduced form models, usable almost only for data description and forecasting. Boivin (2001) considers the opposite case of time varying simultaneous relations, but neglects the potential heteroscedasticity of the innovations. Ciccarelli and Rebucci (2003) extend the framework of Boivin (2001) allowing for t-distributed errors, which account for non-persistent changes in the scale of the variances over time. Uhlig (1997) introduces unrestricted multivariate stochastic volatility in the context of VARs, but his model assumes that the VAR coefficients are constant. Here instead, both the coefficients and the entire variance covariance matrix of the shocks are allowed to vary over time. This is crucial if the objective is distinguishing

between changes in the typical size of the exogenous innovations and changes in the transmission mechanism.

There is also a more recent literature that models time variation in linear structures with discrete breaks, meant to capture a finite number of switching regimes (see, for instance, Hamilton (1989), Kim and Nelson (1999), Sims (1999, 2001*a*), Sims and Zha (2004)). Discrete breaks models may well describe some of the rapid shifts in policy. However, they seem less suitable for capturing changes in private sector behaviour, where aggregation among agents usually plays the role of smoothing most of the changes. Furthermore, even in a structural VAR, the private sector equations can be considered as reduced form relations with respect to a possible underlying behavioural model, where policy and private sector behaviour are not easily distinguishable. If policy responds also to expectational future variables (instead of only to current and past ones), then also the policy equation in the VAR will be a mixture of policy and private sector behaviour, determining smoother changes of the coefficients. Finally, the existence of any type of learning dynamics of private agents or the monetary authorities definitely favours a model with smooth and continuous drifting coefficients over a model with discrete breaks.

From the perspective of the empirical application, this paper is related to a large literature that analyses changes in the conduct of monetary policy and their effect on the rest of the economy. Most of the existing academic work has emphasized the role of monetary policy in the poor economic performance of the 1970's (see, among others, Judd and Rudebusch (1998), Clarida *et al*. (2000), Boivin (2001), Cogley and Sargent (2001, 2003), Boivin and Giannoni (2003), Favero and Rovelli (2003), Lubik and Schorfheide (2004)). This paper contrasts the most popular view and stresses the role of heteroscedastic non-policy innovations. In this respect, the conclusions are more similar to Bernanke and Mihov (1998) and Sims and Zha (2004).

The paper is organized as follows. Section 2 presents the time varying structural VAR model adopted in the paper. Section 3 illustrates the key steps of the estimation methodology. Section 4 discusses the empirical results of the application to the U.S. economy. Section 5 concludes.

2. THE MODEL

The model presented in this paper is a multivariate time series model with both time varying coefficients and time varying variance covariance matrix of the additive innovations. The drifting coefficients are meant to capture possible nonlinearities or time variation in the lag structure of the model. The multivariate stochastic volatility is meant to capture possible heteroscedasticity of the shocks and nonlinearities in the simultaneous relations among the variables of the model. Allowing for time variation both in the coefficients and the variance covariance matrix leaves it up to the data to determine whether the time variation of the linear structure derives from changes in the size of the shocks (impulse) or from changes in the propagation mechanism (response).

It is worth noting that the model admits many types of shocks. Nevertheless the heteroscedasticity assumption is limited to the additive innovations. This is not only for reasons of convenience, but also because the time variation in the size of the additive shocks is a feature of many empirical applications in macroeconomics.[1] Furthermore, as recently emphasized by Cogley and Sargent (2003), overlooking heteroscedasticity would generate fictitious dynamics in the random coefficients.

Consider the model

$$y_t = c_t + B_{1,t} y_{t-1} + \cdots + B_{k,t} y_{t-k} + u_t \qquad t = 1, \ldots, T. \tag{1}$$

1. See for instance Bernanke and Mihov (1998) or Sims and Zha (2004).

y_t is an $n \times 1$ vector of observed endogenous variables; c_t is an $n \times 1$ vector of time varying coefficients that multiply constant terms; $B_{i,t}$, $i = 1, \ldots, k$, are $n \times n$ matrices of time varying coefficients; u_t are heteroscedastic unobservable shocks with variance covariance matrix Ω_t. Without loss of generality, consider the triangular reduction of Ω_t, defined by

$$A_t \Omega_t A_t' = \Sigma_t \Sigma_t', \tag{2}$$

where A_t is the lower triangular matrix

$$A_t = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \alpha_{21,t} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \alpha_{n1,t} & \cdots & \alpha_{nn-1,t} & 1 \end{bmatrix}$$

and Σ_t is the diagonal matrix

$$\Sigma_t = \begin{bmatrix} \sigma_{1,t} & 0 & \cdots & 0 \\ 0 & \sigma_{2,t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{n,t} \end{bmatrix}.$$

It follows that

$$\begin{aligned} y_t &= c_t + B_{1,t} y_{t-1} + \cdots + B_{k,t} y_{t-k} + A_t^{-1} \Sigma_t \varepsilon_t, \\ V(\varepsilon_t) &= I_n. \end{aligned} \tag{3}$$

Stacking in a vector B_t all the R.H.S. coefficients, (3) can be rewritten as

$$\begin{aligned} y_t &= X_t' B_t + A_t^{-1} \Sigma_t \varepsilon_t, \\ X_t' &= I_n \otimes [1, y_{t-1}', \ldots, y_{t-k}'], \end{aligned} \tag{4}$$

where the symbol $\otimes$ denotes the Kronecker product.

A decomposition of the variance covariance matrix resulting in (4) is common, especially in the literature considering the problem of efficiently estimating covariance matrices (see, for instance, Pinheiro and Bates (1996), Pourahmadi (1999, 2000), Smith and Kohn (2002)). In the context of time varying VAR models, Cogley (2003) and Cogley and Sargent (2003) have a similar decomposition, but with a time invariant A_t matrix. It is important to notice that allowing the matrix A_t to vary over time is crucial for a time varying structural VAR. A constant A_t would imply that an innovation to the i-th variable has a time invariant effect on the j-th variable. This is clearly undesirable if the objective is modelling time variation in a simultaneous equation model, where simultaneous interactions among variables are fundamental.

The modelling strategy consists of modelling the coefficient processes in (4) instead of (1). Observe that there is a one to one mapping between (1) and (4) that fully justifies this approach. Let α_t be the vector of non-zero and non-one elements of the matrix A_t (stacked by rows) and σ_t be the vector of the diagonal elements of the matrix Σ_t. The dynamics of the model's time varying parameters is specified as follows:

$$B_t = B_{t-1} + \nu_t, \tag{5}$$

$$\alpha_t = \alpha_{t-1} + \zeta_t, \tag{6}$$

$$\log \sigma_t = \log \sigma_{t-1} + \eta_t, \tag{7}$$

where the distributional assumptions as regards $(\varepsilon_t, \nu_t, \zeta_t, \eta_t)$ are stated below. The elements of the vector B_t are modelled as random walks, as are the free elements of the matrix A_t. The standard deviations (σ_t) are assumed to evolve as geometric random walks, belonging to the class of models known as stochastic volatility. This constitutes an alternative to ARCH models. The crucial difference is that the variances generated by (7) are unobservable components.[2]

It is well known that a random walk process hits any upper or lower bound with probability one and this is clearly an undesirable implication of this model. On the other hand, as long as (5), (6) and (7) are thought to be in place for a finite period of time and not forever, this set of assumptions should be innocuous. Moreover, the random walk assumption presents the advantages of focusing on permanent shifts and reducing the number of parameters in the estimation procedure. Notice that, in principle, the model can be easily extended to consider more general autoregressive processes.[3] Especially for the parameters of the variance covariance matrix, this constitutes an advantage of this model over the so-called local scale models (see, for example, Shephard, 1994*a* and the multivariate generalization of Uhlig, 1997).

All the innovations in the model are assumed to be jointly normally distributed with the following assumptions on the variance covariance matrix:

$$V = \text{Var}\left(\begin{bmatrix} \varepsilon_t \\ \nu_t \\ \zeta_t \\ \eta_t \end{bmatrix}\right) = \begin{bmatrix} I_n & 0 & 0 & 0 \\ 0 & Q & 0 & 0 \\ 0 & 0 & S & 0 \\ 0 & 0 & 0 & W \end{bmatrix}, \tag{8}$$

where I_n is an n-dimensional identity matrix, Q, S and W are positive definite matrices. It is worth noting that none of the restrictions on the structure of V are essential. All the zero blocks could be replaced by non-zero blocks, with only small modifications of the estimation procedure that will be described in the next section. Nevertheless, there are at least two factors suggesting a choice of V such as the one described in (8). The first one is related to the already high number of parameters of the model. Adding all the off-diagonal elements of V would require the specification of a sensible prior, able to prevent cases of ill-determined parameters. The second (and more important) reason is that allowing for a completely generic correlation structure among different sources of uncertainty would preclude any structural interpretation of the innovations. A complete discussion of the issue is postponed to the next section.

Most of the paper will adopt the additional assumption of S being block diagonal, with blocks corresponding to parameters belonging to separate equations. In other words, the coefficients of the contemporaneous relations among variables are assumed to evolve independently in each equation. Even though it is not crucial, this assumption simplifies the inference and increases the efficiency of the estimation algorithm. The general case of S unrestricted will be considered in Section 4.4, which shows the robustness of the empirical results to this simplifying assumption.

3. BAYESIAN INFERENCE

The following notation is used to denote the history of a generic vector of variables ω_t up to a generic time τ:

$$\omega^{\tau} = [\omega_1', \ldots, \omega_{\tau}']'.$$

2. See Shephard (1996) for an overview of the econometric properties of the stochastic volatility model and a comparison with ARCH.

3. Section 4.4 will provide some further discussion and evidence.

For a generic matrix of variables and constant terms M_t,

$$M^\tau = [m_1', \ldots, m_\tau']',$$

where m_t is a column vector constructed with the time varying elements of M_t.

The objective of this section is to lay out the econometric techniques used to estimate the time varying structural VAR model presented in the previous section. Bayesian methods are used to evaluate the posterior distributions of the parameters of interest, *i.e.* the unobservable states, B^T, A^T, Σ^T and the hyperparameters of the variance covariance matrix V. Observe that, dealing with unobservable components, where the distinction between parameters and shocks is less clear than in other situations, a Bayesian approach is the natural one. Three other factors make Bayesian methods particularly suitable for estimating this class of models and preferable to classical estimation. First, if the variance of the time varying coefficients is small, the classical maximum likelihood estimator of this variance has a point mass at zero. This is related to the so-called pile-up problem (see, for instance, Sargan and Bhargava (1983), Shephard and Harvey (1990), Stock and Watson (1998)). The second drawback of classical maximum likelihood is related to the high dimensionality and nonlinearity of the problem. Such a complicated model will quite possibly have a likelihood with multiple peaks, some of which are in uninteresting or implausible regions of the parameter space. Moreover, if these peaks are very narrow, the likelihood may reach particularly high values, not at all representative of the model's fit on a wider and more interesting parameter region. In a Bayesian setting, the use of uninformative priors on reasonable regions of the parameter space is nevertheless effective in ruling out these misbehaviours. The third reason is practical: even though it is in principle possible to write up the likelihood of the model, it is a hard task to maximize it over such a high dimensional space. Bayesian methods deal efficiently with the high dimension of the parameter space and the nonlinearities of the model, splitting the original estimation problem in smaller and simpler ones. Here, Gibbs sampling is used for the posterior numerical evaluation of the parameters of interest. Gibbs sampling is a particular variant of Markov chain Monte Carlo (MCMC) methods that consists of drawing from lower dimensional conditional posteriors as opposed to the high dimensional joint posterior of the whole parameter set.

Finally, observe that MCMC is a smoothing method and therefore delivers smoothed estimates, *i.e.* estimates of the parameters of interest based on the entire available set of data. The suitability of smoothed estimates, as opposed to filtered ones, cannot be established *a priori*, but clearly depends on the specific problem at hand. For example, filtered estimates are more appropriate if the objective of the investigation is constructing model diagnostics (and marginal likelihoods) or forecasting evaluation. The use of simulation based filters is required in these cases. Simulation based filters for these kinds of problems are known as particle filters, which are beyond the scope of this paper (for an overview, see Doucet, De Freitas and Gordon, 2001). On the other hand, smoothed estimates are more efficient (and therefore preferable) when, like in this paper, the objective is an investigation of the true evolution of the unobservable states over time. In these cases, filtered estimates are inappropriate because they would exhibit transient variation even in time invariant models (as pointed out by Sims, 2001*b*).

The rest of the section contains a description of the priors, a sketch of the estimation strategy (the details are in Appendix A) and an explanation of how to deal with non-triangular identification structures.

3.1. Priors and ordering

The prior distributions proposed in this paper are chosen because of their intuitiveness and convenience in the applications. First of all, it is convenient to assume that the initial states

for the coefficients, for the covariances, for the log volatilities and the hyperparameters are independent of each other. The priors for the hyperparameters, Q, W and the blocks of S, are assumed to be distributed as independent inverse-Wishart. The priors for the initial states of the time varying coefficients, simultaneous relations and log standard errors, $p(B_0)$, $p(\alpha_0)$ and $p(\log \sigma_0)$, are assumed to be normally distributed.[4] These assumptions together with (5), (6) and (7) imply normal priors on the entire sequences of the B's, α's and $\log \sigma$'s (conditional on Q, W and S).

The normal prior on B is standard. Smith and Kohn (2002) use the same decomposition as in (2) and place a similar, normal prior on the elements of A. In the context of VAR models, Sims and Zha (1998) also use a similar prior. Placing a prior on the elements of A, as opposed to A^{-1}, might seem strange. However, the statistics literature has long recognized the advantages of modelling the inverse of the variance covariance matrix, as opposed to the covariance matrix itself (Dempster (1972), Cox and Wermuth (1996)). The log-normal prior on the σ's is common in the stochastic volatility literature (see, for instance, Harvey *et al*. (1994), Kim *et al*. (1998)). Such a prior is not conjugate, but has the advantage of maintaining tractability. On the other hand, inverse-gamma and inverse-Wishart priors (which are conjugate) have been used for local scale models (Shephard (1994*a*), Uhlig (1997)). See Shephard (1994*b*) for a comparison of the two approaches in the univariate framework.

Finally, it is important to highlight a potential drawback of the modelling assumptions (4), (6), (7) and (8), in combination with the priors on the A's and Σ's. The drawback is the fact that, in principle, the order of the variables matters in (4). This is due to the lower triangular structure imposed on the A_t matrix. More precisely, consider the vector $\tilde{y}_t$, obtained from a permutation of the order of the elements of the vector y_t. Similarly, let $\tilde{\Omega}_t$ denote the covariance matrix obtained from the same permutation of the rows and columns of Ω_t. It can be shown that it is impossible to find a lower triangular matrix $\tilde{A}_t$ and a diagonal matrix $\tilde{\Sigma}_t$ such that

$$\tilde{A}_t^{-1} \tilde{\Sigma}_t \tilde{\Sigma}_t' \tilde{A}_t^{-1\prime} = \tilde{\Omega}_t^*,$$

where the free elements of $\tilde{A}_t$ are normal, the free elements of $\tilde{\Sigma}_t$ are log-normal and $\tilde{\Omega}_t^*$ has the same distribution as $\tilde{\Omega}_t$. One way to see it is by noticing that the $(1, 1)$ element of $\tilde{\Omega}_t^*$ would necessarily be log-normally distributed, while the $(1, 1)$ element of $\tilde{\Omega}_t$ is not.[5] Usually, in a time invariant setting, the lower triangular structure imposed on A does not affect inference (because, for the variance covariance matrix, the likelihood soon dominates the prior). In this time varying setting however, it is less clear what may happen and the relevance of this issue might vary from case to case.

Observe that, if one is particularly concerned about this problem, there exists a natural solution, which is to impose a prior on all plausible orders of the variables of the system. Consequently, results obtained with different orders can be averaged on the basis of the prior or the posterior probabilities of different models (the models' posterior probabilities can be obtained, for instance, using the reversible jump MCMC method, described below). It is worth pointing out that, in the context of the empirical application of this paper, the results obtained with different orders of the variables are very similar.

4. The values of means, variances, degrees of freedom and scale matrices are specified in the next section, where the application to the U.S. data is discussed.

5. For example, in the two-dimensional case, the $(1, 1)$ element of $\tilde{\Omega}_t$ would be the sum of a log-normal distribution and the product of another log-normal distribution and a normal distribution squared. This sum is not log-normal.

3.2. *Simulation method*

The model is estimated by simulating the distribution of the parameters of interest, given the data. While the details are left to Appendix A, this section sketches the MCMC algorithm used to generate a sample from the joint posterior of $\left(B^T, A^T, \Sigma^T, V\right)$. As mentioned above, Gibbs sampling is used in order to exploit the blocking structure of the unknowns. Gibbs sampling is carried out in four steps, drawing in turn time varying coefficients (B^T), simultaneous relations (A^T), volatilities (Σ^T) and hyperparameters (V), conditional on the observed data and the rest of the parameters.

Conditional on A^T and Σ^T, the state space form given by (4) and (5) is linear and Gaussian. Therefore, the conditional posterior of B^T is a product of Gaussian densities and B^T can be drawn using a standard simulation smoother (Carter and Kohn (1994) or Fruhwirth-Schnatter (1994)). For the same reason, the posterior of A^T conditional on B^T and Σ^T is also a product of normal distributions. Hence A^T can be drawn in the same way. Drawing Σ^T instead is more involved and relies mostly on the method presented in Kim *et al*. (1998). It consists of transforming a nonlinear and non-Gaussian state space form in a linear and approximately Gaussian one, which, again, allows the use of standard simulation smoothers. Simulating the conditional posterior of V is standard, since it is the product of independent inverse-Wishart distributions.

Observe that the model and the posterior simulation method are close to Shephard (1994*b*). Indeed, the model belongs to the abstract class of partially non-Gaussian state space models, introduced by Shephard (1994*b*). However, this abstract class of models is quite broad and in Shephard (1994*b*) there is no treatment for the general multivariate case.

3.3. *Identification and structural interpretation*

The method described in the previous subsection and Appendix A allows one to estimate a reduced form VAR. However, as long as an exact identification scheme for the additive shocks is available, a structural VAR can be easily estimated in two steps. Consider the following structural VAR:

$$y_t = X_t' B_t + \Xi_t \varepsilon_t,$$

that differs from (4) because the $n \times n$ matrices Ξ_t, $t = 1, \ldots, T$, are not necessarily lower triangular. Assume further that for any t, Ξ_t contains at least $\frac{n(n-1)}{2}$ restrictions which guarantees identification.[6] The first step consists of estimating the reduced form VAR, following the methodology illustrated in the previous sections. This first step delivers the posterior of the B's and the Ω's[7] at every point in time. To obtain the numerical evaluation of the posterior of the Ξ's, it suffices to solve the system of equations given by

$$\Xi_t \Xi_t' = \Omega_t, \qquad t = 1, \ldots, T \tag{9}$$

for every draw of Ω_t. Observe that, even though this is a fairly general procedure, there may exist identifying structures of Ξ_t for which (9) has no solution. As an illustrative example, consider the following structure for Ξ_t, in a three-by-three case:

$$\Xi_t = \begin{bmatrix} x & 0 & 0 \\ 0 & x & x \\ x & x & x \end{bmatrix}, \tag{10}$$

6. These can be zero restrictions or cross-element restrictions.
7. Remember that Ω_t, $t = 1, \ldots, T$, are the variance covariance matrices of the errors of the reduced form VAR at every point in time.

where the x's indicate the elements potentially different from zero. Equation (10) implies a variance covariance matrix of the reduced form residuals with zero in the position $(1, 2)$ that would not be compatible with draws of Ω_t obtained using the decomposition in (4) unless the element $(1, 2)$ of the draw is very close to zero (in this case the solution would be approximate, but probably still reliable). If the draw of Ω_t has the $(1, 2)$ element far from zero, the draw must be discarded. It is clear that the method is not efficient if the rejection rate is high. On the other hand, this is not too worrying if a high rejection rate is thought of as a signal of a structure for Ξ_t rejected by the data.

Clearly, if the identification is based on a triangular scheme (like the one in the next section), the solution to (9) is simply given by $\Xi_t = A_t^{-1}\Sigma_t$. Observe that this framework does not allow an efficient estimation of overidentified systems, unless the overidentification derives from further zero restrictions in a triangular scheme. This case is easily made tractable by imposing the corresponding restrictions.

As anticipated in Section 2, this is the right framework for attempting some structural interpretations. The fact that the elements of Ξ are time varying represents the crucial difference between modelling time variation in a structural VAR as opposed to a standard VAR. In fact, models characterized by a time invariant variance covariance matrix (Canova (1993), Sims (1993) or Cogley and Sargent (2001)) or in which only the variances are drifting (Cogley (2003), Cogley and Sargent (2003)) imply that all the contemporaneous relations among variables are time invariant. And this seems undesirable in a context in which all the lag coefficients are time varying. The other extreme, a model with time varying variance covariance matrix but constant coefficients (Uhlig, 1997) is probably less undesirable but may appear restrictive too.

Of course, the flexibility of the model of this paper does not come without costs. And the main one derives exactly from the lack of strong restrictions, able to isolate the sources of uncertainty. An example will clarify the issue: consider the common specification of the "monetary policy" VAR used below (a three-variable VAR with inflation, unemployment and a short-term interest rate). Assume that a reliable identification scheme is available for the monetary policy equation (say a triangular identification based on the assumptions that inflation and unemployment react to the policy instrument with at least one period of lag). In a time invariant framework this would allow the use of standard structural VAR tools, as impulse responses and variance decompositions. This is because there is just one source of shocks for the monetary policy equation, that is the usual additive innovation. The situation is not as simple for a time varying structural VAR because the potential sources of monetary policy shocks are many and depend on the specific assumptions on the variance covariance matrix of all the innovations of the model. Looking at equation (8) for example, in the model described in this paper there would be at least three independent sources of monetary policy shocks.[8] The first is the usual additive shock (ε). The second is a shock to the interest rate reaction to *current* inflation and unemployment (ζ). The third is a shock to the interest rate reaction to *past* values of the variables of the system (ν). Monetary policy shocks of the first type are assumed to be independent of any other innovation. This assumption is crucial for the interpretation of the effects of such shocks, besides the existence of other innovations in the system. On the other hand, shocks of the second type are assumed to be correlated within equation, but uncorrelated across equations (if S is assumed block diagonal). This assumption would allow, for example, the analysis of the impact of shocks to the systematic part of monetary policy on the rest of the economy. Shocks of the third type are instead potentially correlated also across equations. This means that innovations of the third type to the non-policy equations will potentially affect also the monetary policy

8. Not considering the shocks to the standard deviations of the additive innovations.

equation. However, this is not necessarily a bad assumption. On the contrary, it seems to move in the direction of more realism in admitting that there exist sources of uncertainty in the system that cannot be easily separable and othogonalized for each equation. Actually, a full correlation structure for the block $[\zeta', \nu']$ would be even more appropriate, particularly if the VAR is thought as a reduced form representation of a forward looking model, where most of the coefficients of the VAR would be convolutions of the original parameters of the model. However, while it is relatively easy to allow for a full variance covariance matrix for the block $[\zeta', \nu']$, it would come at the cost of a substantially increased number of parameters. Section 4.4 will provide an evaluation of the empirical importance of cross-equation correlation among parameters, showing that the model is not restrictive in practice. An additional note of caution is necessary: in fact, if a full correlation structure among parameters exists, it would probably show up in a dynamic setting, especially if learning plays a role in the economy. This potential kind of correlation is absent in the model.

4. A SMALL MODEL OF THE U.S. ECONOMY

The techniques just described are applied for the estimation of a small quarterly model of the U.S. economy. Three variables are included in the model: inflation rate and unemployment rate, representing the non-policy block, and a short-term nominal interest rate, representing the policy block.[9] Even though most of the literature has preferred larger sets of variables (see, for instance, Leeper, Sims and Zha (1996), Bernanke and Mihov (1998), Christiano, Eichenbaum and Evans (1999)), there are a few existing papers that use the same small system of equations (for example, Rotemberg and Woodford (1997), Cogley and Sargent (2001, 2003), Stock and Watson (2001)). Here, the choice is mostly due to the attempt to reduce the number of parameters of the model. As already mentioned, the alternative would be to use a wider set of variables at the cost of tighter priors, necessary to avoid ill behaviours.[10]

The sample runs from 1953:I to 2001:III. Two lags are used for the estimation. The simulations are based on 10,000 iterations of the Gibbs sampler, discarding the first 2000 for convergence. As shown in Appendix B, the sample autocorrelation functions of the draws decay quite fast and the convergence checks are satisfactory.

4.1. Priors

The first 10 years (40 observations, from 1953:I to 1962:IV) are used to calibrate the prior distributions. For example, the mean and the variance of B_0 are chosen to be the *OLS* point estimates ($\widehat{B}_{OLS}$) and four times its variance in a time invariant VAR, estimated on the small initial subsample. In the same way, a reasonable prior for A_0 can be obtained. For $\log \sigma_0$ instead, the mean of the distribution is chosen to be the logarithm of the *OLS* point estimates of the standard errors of the same time invariant VAR, while the variance covariance matrix is arbitrarily

9. All series are taken from the Standard and Poor's DRI database. Inflation is measured by the annual growth rate of a chain weighted GDP price index. Unemployment is the unemployment rate referred to all workers over 16. The nominal interest rate is the yield on three-month Treasury bills, preferred to the more conventional federal fund rate because they are available for a longer period of time.

10. In order to solve this problem, an interesting line for future research would be to assume the existence of common factors driving the dynamics of the coefficients. This approach would be very useful to reduce the rank of the variance covariance matrix of the innovations to the coefficients on the lags. The idea is the following: replace (5) with $B_t = B_{t-1} + C\xi_t$, where ξ_t has dimension lower than B_t. This can be handled by the estimation procedure of Section 3, adding an extra step in order to draw the C matrix. The state equation can be rewritten as $B_t = B_0 + C \sum_{s=1}^{t} \xi_s$ and substituted in the measurement equation (4). Conditional on the rest of the parameters, the posterior of $vec(C)$ turns out to be normally distributed with known mean and variance.

assumed to be the identity matrix. Finally, degrees of freedom and scale matrices are needed for the inverse-Wishart prior distributions of the hyperparameters. The degrees of freedom are set to 4 for W and 2 and 3 for the two blocks of S (basically, one plus the dimension of each matrix). The reason that the degrees of freedom are chosen differently is that for the inverse-Wishart distribution to be proper the degrees of freedom must exceed the dimension respectively to W and the blocks of S. For Q the degrees of freedom are set to 40 (the size of the previous initial subsample), since a slightly tighter prior seems to be necessary in order to avoid implausible behaviours of the time varying coefficients (see Section 4.4 for a detailed discussion). Following the literature (Cogley and Sargent (2001, 2003), Cogley (2003)), the scale matrices, $\overline{Q}$, $\overline{W}$, $\overline{S}_1$ and $\overline{S}_2$, are chosen to be constant fractions of the variances of the corresponding *OLS* estimates on the initial subsample (multiplied by the degrees of freedom, because, in the inverse-Wishart distribution, the scale matrix has the interpretation of sum of squared residuals). Summarizing, the priors take the forms:

$$\begin{aligned}
B_0 &\sim N(\widehat{B}_{OLS}, 4 \cdot V(\widehat{B}_{OLS})),\\
A_0 &\sim N(\widehat{A}_{OLS}, 4 \cdot V(\widehat{A}_{OLS})),\\
\log \sigma_0 &\sim N(\log \widehat{\sigma}_{OLS}, I_n),\\
Q &\sim IW(k_Q^2 \cdot 40 \cdot V(\widehat{B}_{OLS}), 40),\\
W &\sim IW(k_W^2 \cdot 4 \cdot I_n, 4),\\
S_1 &\sim IW(k_S^2 \cdot 2 \cdot V(\widehat{A}_{1,OLS}), 2),\\
S_2 &\sim IW(k_S^2 \cdot 3 \cdot V(\widehat{A}_{2,OLS}), 3),
\end{aligned}$$

where S_1 and S_2 denote the two blocks of S, while $\widehat{A}_{1,OLS}$ and $\widehat{A}_{2,OLS}$ stand for the two corresponding blocks of $\widehat{A}_{OLS}$. The benchmark results presented in this section are obtained using the following values: $k_Q = 0{\cdot}01$, $k_S = 0{\cdot}1$, $k_W = 0{\cdot}01$. Set in this way, the priors are not flat, but diffuse and uninformative. The reader is referred to Section 4.4 for a detailed discussion of this choice and of the robustness of the results to alternative prior specifications. The discussion of the empirical results starts with the analysis of time variation of the U.S. non-systematic monetary policy.

4.2. Non-systematic monetary policy

The term non-systematic monetary policy is used to capture both "policy mistakes" and interest rate movements that are responses to variables other than inflation and unemployment (therefore exogenous in this set-up). Identified monetary policy shocks are the measure of non-systematic policy actions. The identifying assumption for the monetary policy shocks is that monetary policy actions affect inflation and unemployment with at least one period of lag. Therefore, interest rates are ordered last in the VAR. It is important to stress that this is not an ordering issue (like the one highlighted in Section 3.1), but an identification condition, essential for isolating monetary policy shocks. Moreover, this identification assumption is completely standard in most of the existing literature (for instance, Leeper *et al*. (1996), Rotemberg and Woodford (1997), Bernanke and Mihov (1998), Christiano *et al*. (1999)). On the other hand, the simultaneous interaction between inflation and unemployment is arbitrarily modelled in a lower triangular form, with inflation first. As opposed to the previous one, this is not an identification condition, but a necessary normalization. While this arbitrary normalization could potentially make a difference (for the reasons illustrated in Section 3.1), in the context of this empirical application, it turns out that the ordering of the non-policy block does not affect the results.

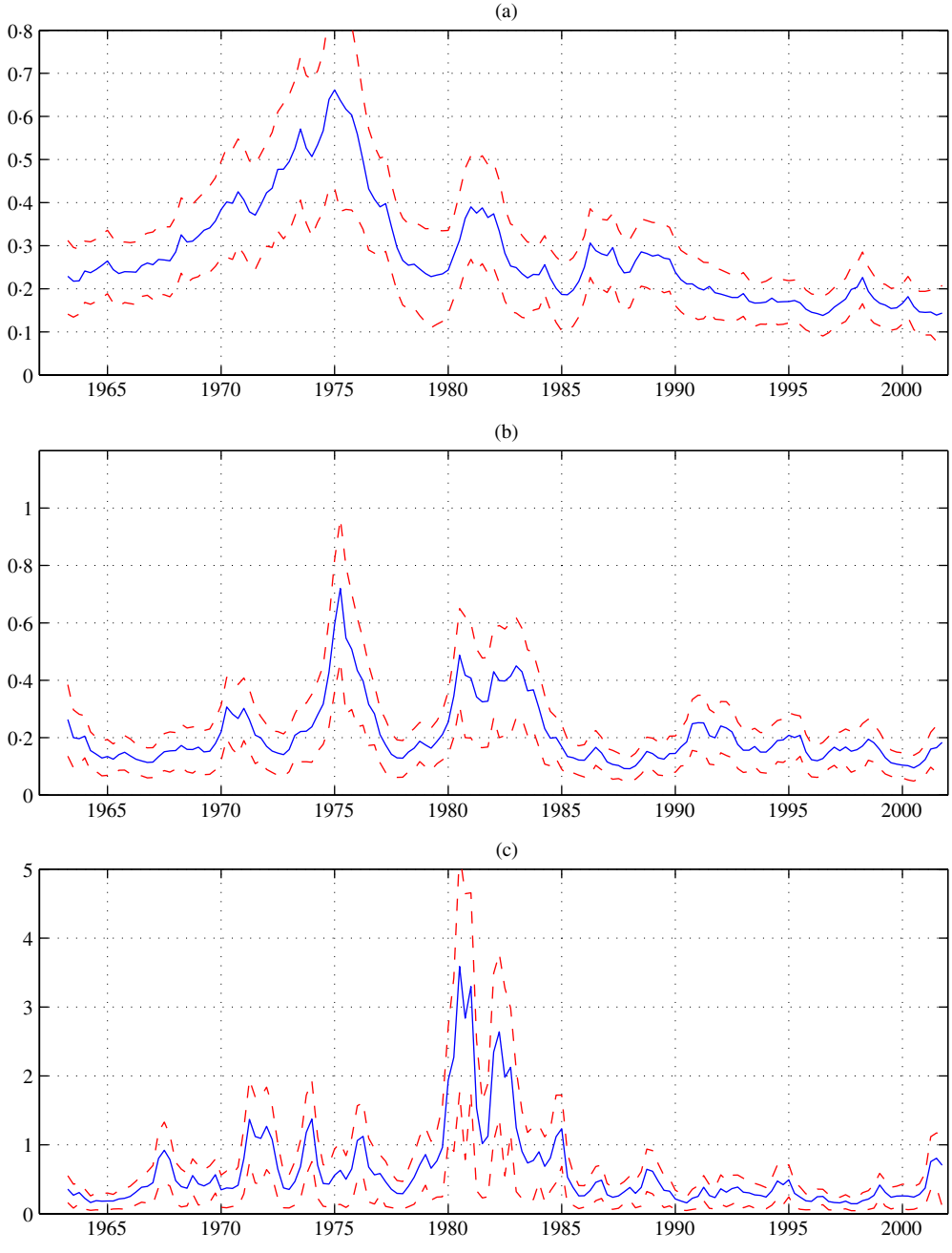

FIGURE 1

Posterior mean, 16-th and 84-th percentiles of the standard deviation of (a) residuals of the inflation equation, (b) residuals of the unemployment equation and (c) residuals of the interest rate equation or monetary policy shocks

If identified monetary policy shocks are the measure of non-systematic policy actions, it seems natural to measure the relative importance and changes of non-systematic monetary policy by the time varying standard deviation of the identified monetary policy shocks. Figure 1(c)

presents a plot of the posterior mean and the 16-th and 84-th percentiles[11] of the time varying standard deviation of the monetary policy shocks. This graph presents at least two interesting features. The period 79–83 exhibits a substantially higher variance of monetary policy shocks. This is not surprising at all and confirms that the Volcker regime and his monetary aggregates targeting was actually peculiar in the history of U.S. monetary policy. What is more interesting is the fact that the volatility of monetary policy shocks was on average higher in the pre-Volcker period than in the post-Volcker one, when it is very low and substantially constant. This suggests that Taylor-type rules (like the one estimated in this paper) are very good approximations of the U.S. monetary policy in the last 15 years, while it is likely that in the 1960's and 1970's the Fed was responding to variables other than just inflation and unemployment.[12] The result that monetary policy shocks were more volatile before 1983 is robust to other specifications of the interest rate equation. One obvious alternative is using the interest rates in logs instead of levels and, consequently, interpreting the coefficients of the interest rate equation as semi-elasticities. This modification of the baseline setting allows the interest rate responses to inflation and real activity to differ, depending on the level of the interest rate itself. In other words, high responses (and, therefore, large interest rate variability) are more likely when interest rates are high and vice versa. This specification of the policy equation has been also suggested by Sims (1999, 2001*a*). The estimation of the model using the log of the interest rate does not change the main characteristics of Figure 1(c). Overall, the result that the volatility of monetary policy shocks is higher in the pre-Volcker period and in the first half of the Volcker chairmanship is robust and consistent with Bernanke and Mihov (1998), Sims (1999, 2001*a*) and Sims and Zha (2004).

The changes in the effects of non-systematic policy are summarized in Figures 2 and 3. Figures 2(a) and 3(a) plot the impulse responses of inflation and unemployment to a monetary policy shock in three different dates of the sample. The other graphs of Figures 2 and 3 represent pairwise differences between impulse responses in different dates with the 16-th and 84-th percentiles. The dates chosen for the comparison are 1975:I, 1981:III and 1996:I. They are somewhat representative of the typical economic conditions of the chairmanships of Burns, Volcker and Greenspan, but, apart from that, they are chosen arbitrarily.[13] Clearly, these responses do not vary much over time, indicating that the estimated coefficients do not show much time variation. Some differences are detectable in the responses of inflation which, in the Burns and Volcker periods, exhibit a small price puzzle. The price puzzle almost disappears in the Greenspan period. However such an effect does not seem to be statistically significant once standard errors are taken into account. Moreover, the stability of the response of unemployment to policy shocks is remarkable.

Summarizing, there is no evidence of nonlinearities in the responses of the economy to non-systematic policy.

4.3. Systematic monetary policy

Common and theoretically important measures of the degree of activism of the systematic monetary policy are the responses of the interest rate to inflation and unemployment. To be

11. Under normality, the 16-th and 84-th percentiles correspond to the bounds of a one-standard-deviation confidence interval.

12. Interestingly, Figure 1 also shows how not only monetary policy shocks, but also the shocks to the inflation and unemployment equations have a lower standard deviation in the second part of the sample.

13. Actually 1975:I is a NBER business cycle trough date, 1981:III is a NBER business cycle peak date and 1996:I is the middle date between the last NBER trough and peak dates (1991:I and 2001:I). They are meant to capture very different economic conditions. Experiments with different dates and with averages over the periods 1970:I–1978:I, 1979:IV–1983:IV and 1987:III–2001:III (Burns chairmanship, Volcker monetary targeting and Greenspan chairmanship) give very similar conclusions.

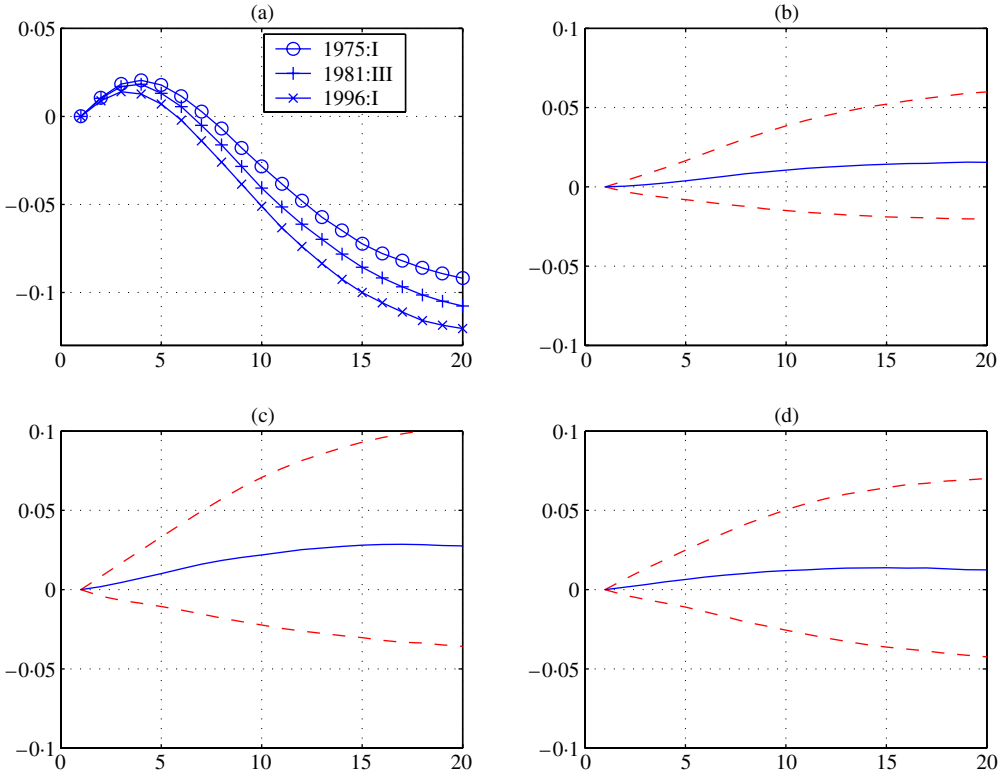

FIGURE 2

(a) Impulse responses of inflation to monetary policy shocks in 1975:I, 1981:III and 1996:I, (b) difference between the responses in 1975:I and 1981:III with 16-th and 84-th percentiles, (c) difference between the responses in 1975:I and 1996:I with 16-th and 84-th percentiles, (d) difference between the responses in 1981:III and 1996:I with 16-th and 84-th percentiles

concrete, consider the modified version of the monetary policy rule of Clarida *et al*. (2000), given by the following equations:

$$i_t = \rho(L) i_{t-1} + (1 - \rho(1)) i_t^* + e_t^{mp} \tag{11}$$

$$i_t^* = i^* + \tilde{\phi}(L) \left(\pi_t - \pi^* \right). \tag{12}$$

i_t is a short-term interest rate; i_t^* is a target for the short-term interest rate; π_t is the inflation rate; π^* is the target for inflation; i^* is the desired interest rate when inflation is at its target; e^{mp} is a monetary policy shock; $\rho(L)$ and $\tilde{\phi}(L)$ are lag polynomials. In other words, the assumption is that the monetary authorities specify a target for the short-term interest rate, but they achieve it gradually (because they tend to smooth interest rate movements). Finally, observe that, just for simplicity, (12) does not allow for any policy response to real activity. The combination of (11) and (12) gives

$$i_t = \bar{i} + \rho(L) i_{t-1} + \phi(L) \pi_t + e_t^{mp}, \tag{13}$$

where $\bar{i} \equiv (1 - \rho(1)) \left(i^* - \tilde{\phi}(1) \pi^* \right)$ and $\phi(L) \equiv (1 - \rho(1)) \tilde{\phi}(L)$. Equation (13) is interpretable as a Taylor rule, augmented in order to consider dynamically richer interest rate responses to inflation. In the context of the popular New-Keynesian framework (see Woodford, 2003

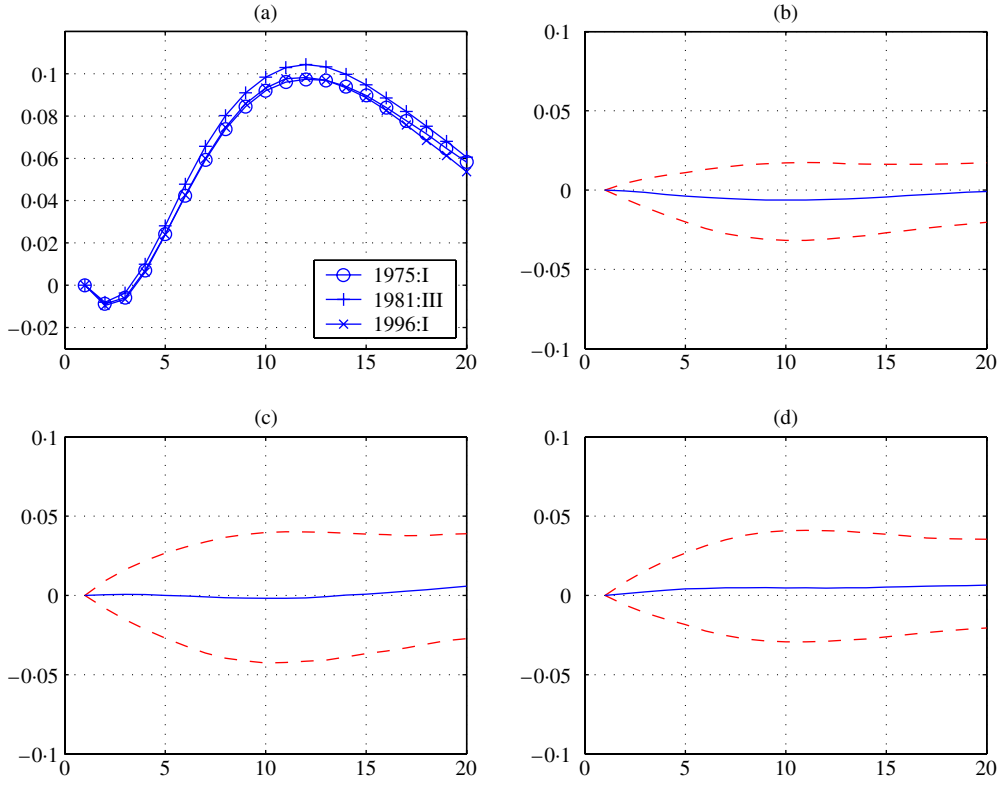

FIGURE 3

(a) Impulse responses of unemployment to monetary policy shocks in 1975:I, 1981:III and 1996:I, (b) difference between the responses in 1975:I and 1981:III with 16-th and 84-th percentiles, (c) difference between the responses in 1975:I and 1996:I with 16-th and 84-th percentiles, (d) difference between the responses in 1981:III and 1996:I with 16-th and 84-th percentiles

for an overview), it has been argued that the Taylor principle is a necessary and sufficient condition under which the rational expectation equilibrium exhibits desirable properties, such as determinacy and expectational stability (Bullard and Mitra (2000), Woodford (2001, 2003)).[14] In this general framework, the Taylor principle requires that

$$\tilde{\phi}(1) > 1,$$

i.e. that the sum of coefficients of the target interest rate response to inflation must be bigger than one. Observe that, in this framework, $\tilde{\phi}(1)$ coincides with (and, consequently, provides an alternative interpretation for) $\frac{\phi(1)}{(1-\rho(1))}$, *i.e.* the long run interest rate response to inflation, implied by the policy rule. Therefore, from a theoretical point of view, long run responses are fundamental. From an empirical point of view, finite time responses of the interest rate to unemployment and inflation seem more informative and realistic quantities. Recall that the time varying variance covariance matrix plays a key role in the analysis of the simultaneous and long run responses that follows.

14. The Taylor principle is only sufficient for determinacy and expectational stability if the monetary policy rule includes an interest rate response to real activity (Woodford, 2001).

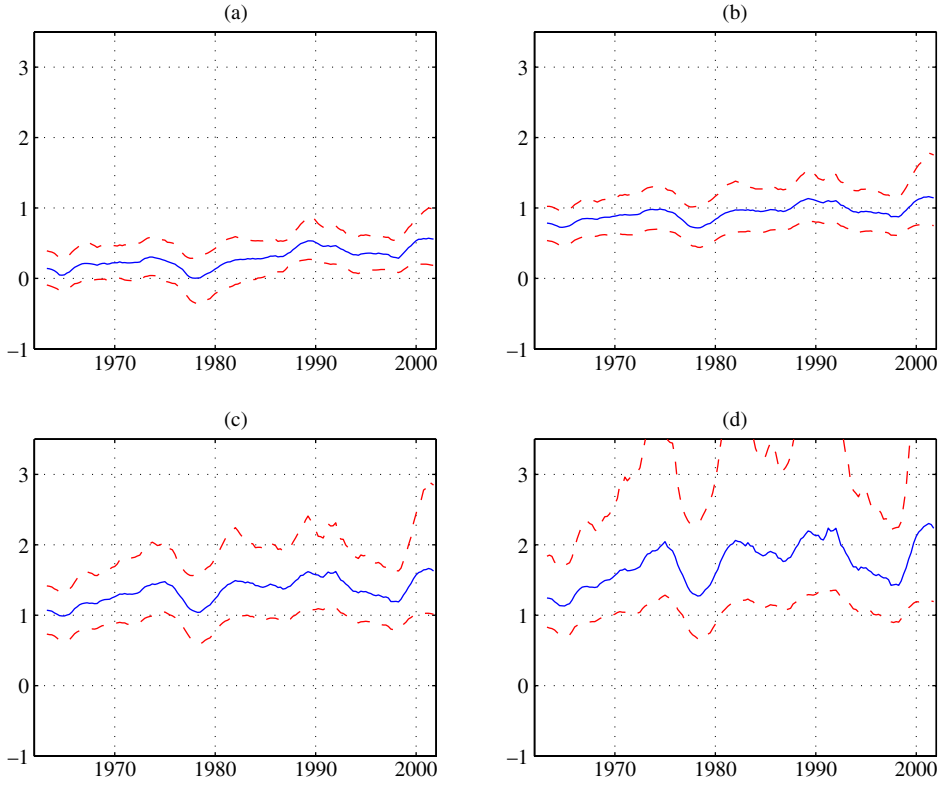

FIGURE 4

Interest rate response to a 1% permanent increase of inflation with 16-th and 84-th percentiles. (a) Simultaneous response, (b) response after 10 quarters, (c) response after 20 quarters, (d) response after 60 quarters

Figures 4 and 6 present plots of the evolution of the responses of interest rate to a 1% permanent increase in inflation and unemployment over the last 40 years. Chart (a) represents the simultaneous response, charts (b), (c) and (d) the responses after 10, 20 and 60 quarters. 16-th and 84-th percentiles are also reported. Even if not exactly the same, the 60-quarter response may be thought as a long run response (or, alternatively, the response of the target interest rate) and therefore can be used to check the Taylor principle.[15]

A few results stand out. First, the correlation between simultaneous and long run responses to inflation is high (Figure 5), even though for the simultaneous response the upward sloping trend is less pronounced. Second, the response of interest rate to inflation is often gradual. In other words, it takes time for the interest rate to reach the long run response level after an inflationary shock. Notice that usually, less aggressive contemporaneous responses reach approximately the same level of the aggressive responses in less than 10 quarters. The dynamic response to unemployment behaves quite differently. As shown in Figure 7, the contemporaneous response

15. Note that the uncertainty around these responses increases with time. In particular, the posterior is clearly skewed toward high level for the response to inflation and low levels for the response to unemployment. This is due to the persistence of the interest rate that implies a sum of coefficients often close to one. This makes the long run responses explode. For the same reason, the graphs report posterior medians instead of posterior means (which would be even higher in absolute value). Finally, notice that the responses do not include the uncertainty about the future evolution of the coefficients.

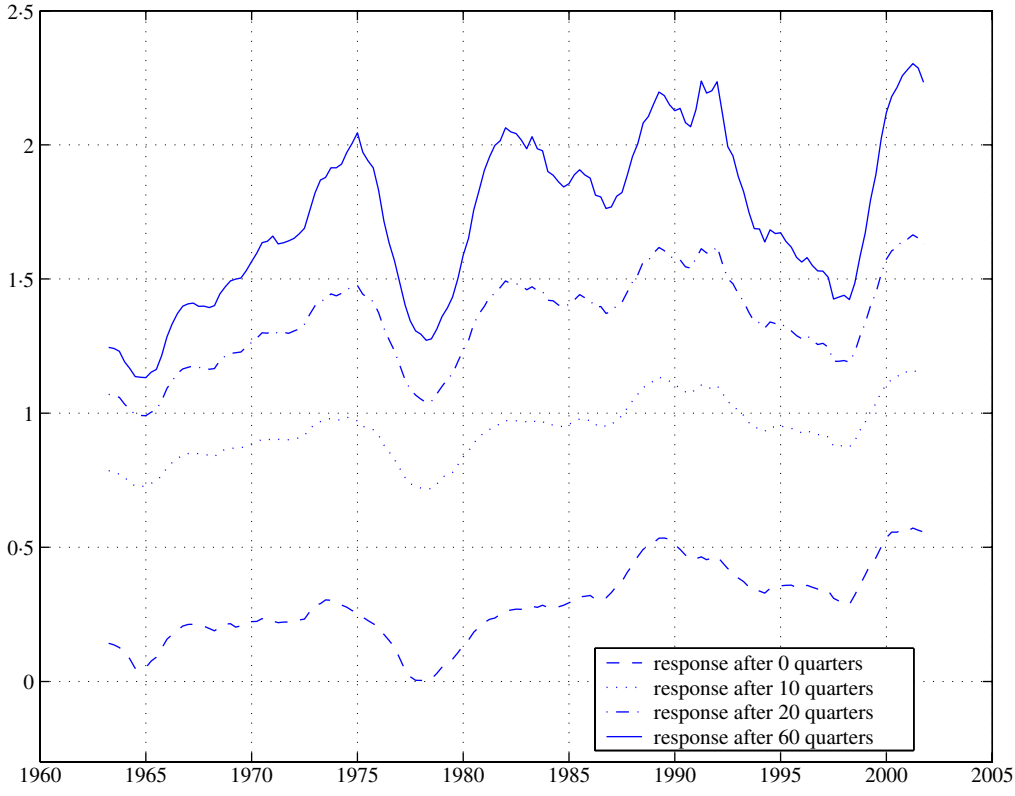

FIGURE 5
Interest rate response to a 1% permanent increase of inflation

is almost the same as the long run one, suggesting that the Fed reacts to unemployment much faster than to inflation. This is not surprising and due to the fact that the signal-to-noise ratio for unemployment is probably perceived to be higher than for inflation. Apart from that, responses to inflation and unemployment are quite highly correlated. This indicates that the periods in which the Fed is more aggressive against inflation are the same periods in which monetary policy is more reactive to unemployment fluctuations.

Third, looking at long run reactions, as many authors have emphasized, systematic monetary policy has become more reactive to inflation and unemployment over the last 20 years with respect to the 1960's and 1970's. However, the 1960's and 1970's do not seem to be characterized by a violation of the Taylor principle. In fact, the long run reaction of the interest rate to inflation remains above one for the whole sample period. This result differs from a large part of the literature that has found an interest rate reaction to inflation lower than one before the beginning of the 1980's (see among others Judd and Rudebusch (1998), Clarida *et al*. (2000), Cogley and Sargent (2001)). This difference can be explained by the fact that this paper presents *smoothed* estimates that use all the information of the sample, as opposed to *filtered* estimates that use only the information contained in the relevant subsample. Strictly related to this, another reason that may drive the conclusions reached by the previous literature is the fact that subsample estimates may suffer from the well-known bias toward stationarity that characterizes many dynamic models estimated on small samples. If for some reason (as it seems to be) the bias toward stationarity is

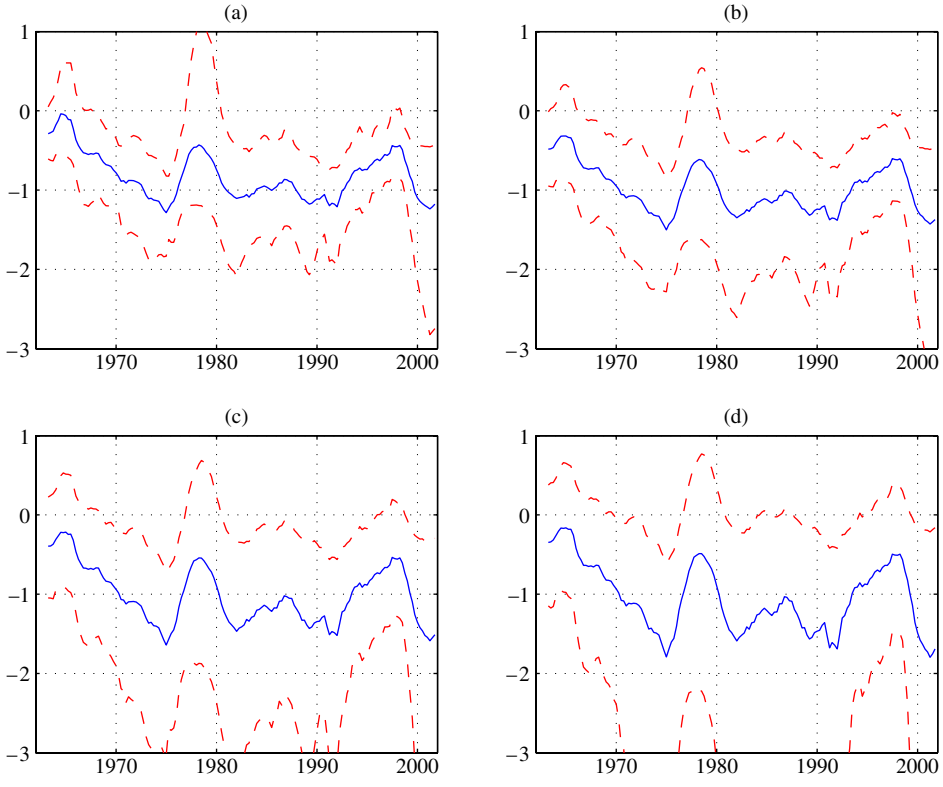

FIGURE 6

Interest rate response to a 1% permanent increase of unemployment with 16-th and 84-th percentiles. (a) Simultaneous response, (b) response after 10 quarters, (c) response after 20 quarters, (d) response after 60 quarters

larger in the subsample of the 1960's and 1970's this creates a larger bias of the long run reactions of interest rate toward low levels.[16]

Fourth, neither the simultaneous reactions nor the long run ones seem to exhibit an absorbing state, but rather a frequent alternation among states. This result partly confirms Sims' (1999; 2001*a*) view of non-unidirectional regime switches, even though the upward sloping trend of the long run response to inflation provides some evidence of a learning pattern toward a more aggressive policy.

In order to establish the responsibility of the Fed in the high inflation and unemployment episodes of the 1970's and early 1980's, it is particularly interesting to analyse the effects of these changes of the systematic part of the policy on the rest of the economy. The methodology for doing so is straightforward: observing the data, all the parameters of the model can be drawn from the joint posterior (as described in Section 3 and Appendix A). For every draw it is possible to reconstruct the i.i.d. sequence of unit-variance structural shocks of the model. Starting from 1970:I these shocks can be used to simulate counterfactual data, constructed using different values of the parameters. These new series can be interpreted as the realization of the data that

16. A way of checking the presence of a strong bias is to estimate a time invariant VAR on the relevant sample and construct a 95% confidence interval based on the implied unconditional distribution. If the initial conditions of the variables lie way outside the interval, this is a signal of the presence of some sort of bias. See Sims (2000) for details.

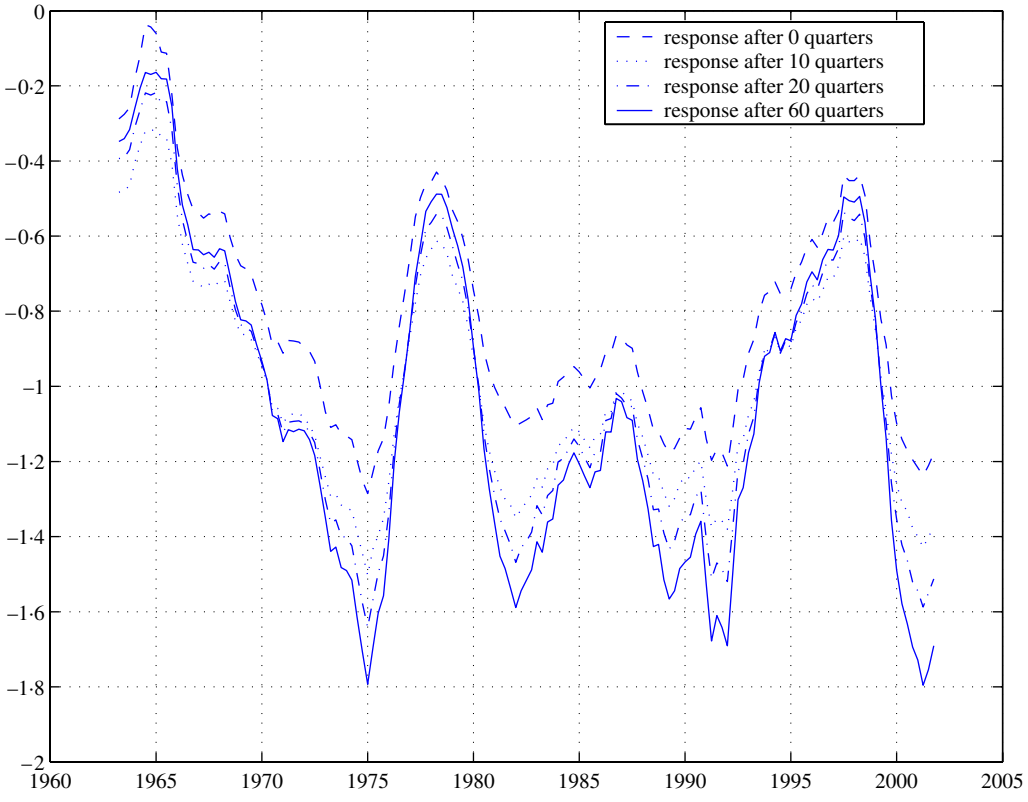

FIGURE 7
Interest rate response to a 1% permanent increase of unemployment

would have been observed had the parameters of the models been the ones used to generate the series. In the context of this paper, the interesting experiment is "planting Greenspan into the 1970's". In other words, the idea consists of replaying history drawing the parameters of the policy rule in the 1970's from their posterior in 1991–1992,[17] in order to see whether this would have made any difference. This is done in Figure 8, from which it is clear that the counterfactual paths of inflation and unemployment do not differ much from the actual ones. Notice that a very similar conclusion is reached if history is replayed using in the 1970's the standard deviation of the monetary policy shocks drawn from its 1991–1992 posterior.

Of course, a Lucas (1976) critique issue arises in this type of counterfactual analysis. If the policy rules adopted in the 1970's and in the 1990's were different, rational and forward looking private agents would take this into account and modify their behaviour accordingly. Therefore, the interpretation of "Greenspan into the 1970's" is not straightforward, since the counterfactual exercise takes the private sector behaviour in the 1970's as given and unchanged. However, there are many reasons that doing the exercise is interesting anyway. First, the effects of the Lucas critique are drastically mitigated in any Bayesian framework, in which policy is random. This argument is even stronger in this framework, in which there is an explicit model of

17. 1991–1992 posterior means the posterior of the average value of the parameters in the eight quarters of 1991 and 1992. Again, the choice of the dates is arbitrary, meant to capture a period of relative economic stability, with relatively strong reactions to unemployment and inflation.

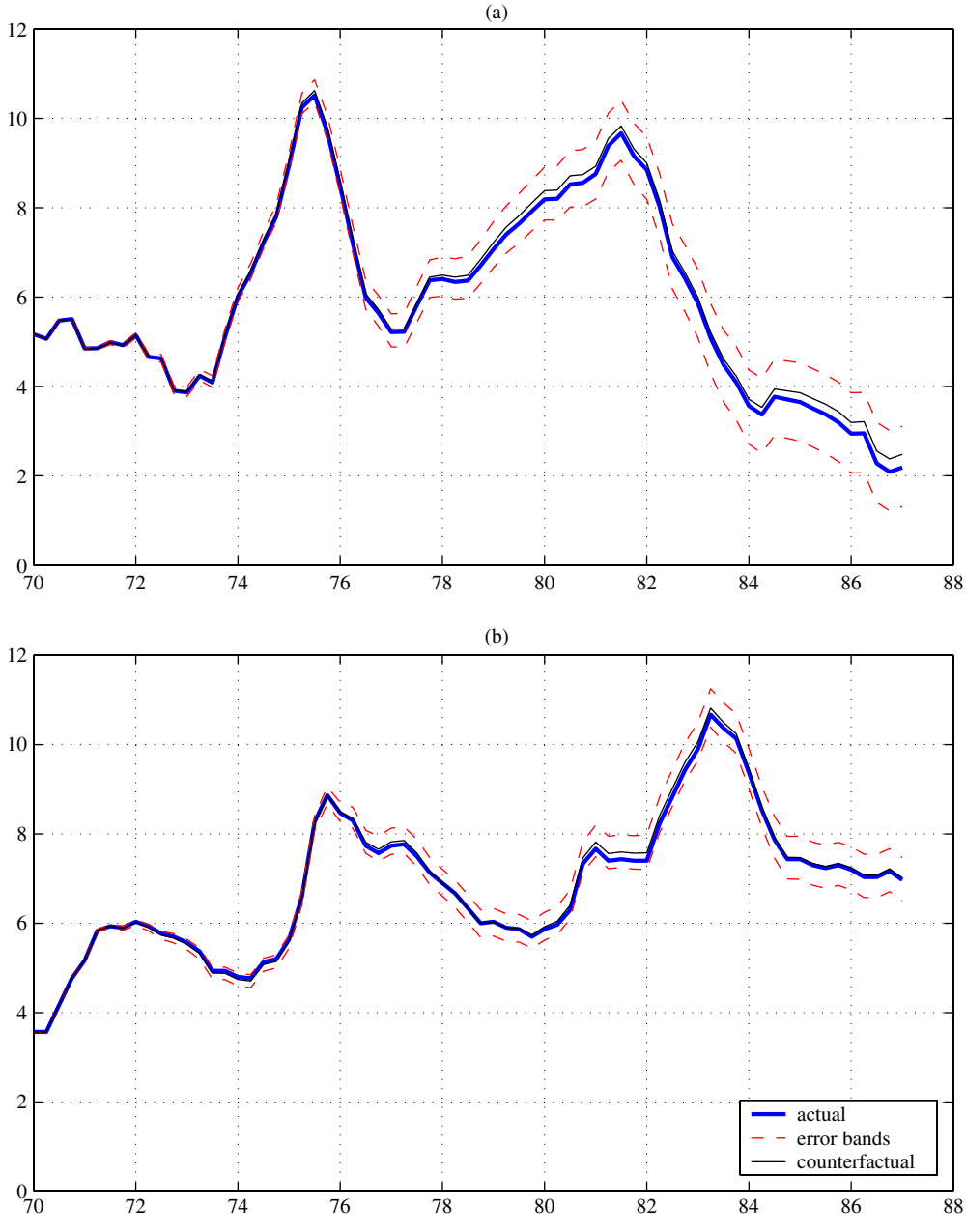

FIGURE 8

Counterfactual historical simulation drawing the parameters of the monetary policy rule from their 1991–1992 posterior. (a) Inflation, (b) unemployment

the stochastic time variation of policy. Second, according to the previous estimates, the changes in policy between the Burns period and the Greenspan period do not seem to be drastic enough to generate a different, previously unknown regime. In other words, the policy parameters of 1991–1992 seem to belong to the private sector probability distribution over the policy parameters for

almost all the 1970–1987 period, as appears from Figures 4 and 6 (even though these figures show the smoothed estimates of the posteriors instead of the filtered ones). Third, the Lucas critique would be a major problem if the counterfactual paths of Figure 8 were very different from the actual paths of inflation and unemployment. Since they are not, it is less clear why rational private agents would modify their behaviour in response to changes of the policy rule, which have minor implications. Fourth, even when the counterfactual experiment is repeated, drawing not only the monetary policy, but also the private sector's parameters from their 1991–1992 posterior, the picture (not reported) looks fundamentally unchanged.

Overall, these findings are in contrast with the anecdotal view that monetary policy was very loose in the 1970's and this caused the poor economic performance of that decade. On the contrary, the impression is that the differences in the conduct of systematic interest rate policies under Burns, Volcker and Greenspan were not large enough to have any relevant effect on the dynamics of inflation and unemployment. Indeed, additional counterfactual simulations (not reported) show that the peaks of inflation and unemployment of the 1970's and early 1980's seem to be better (although not completely) explained by the high volatility of the shocks that characterized those periods. Figure 1(a) and (b) show clearly that inflation and unemployment residuals were less volatile after 1985. Inflation residuals' variability reaches its peak around 1975, oscillating and finally declining since 1982. Less smooth seem to be the changes in the unemployment residuals' variability. In particular, the convergence toward a more stable regime starts with a rapid volatility decrease between 1983 and 1985. These findings are in line with other studies such as Blanchard and Simon (2001), Stock and Watson (2002), Hanson (2003), and Sims and Zha (2004).[18]

Of course, the small model used in this paper is not able to address the issue of the true behavioural source of these sequences of particularly bad shocks. However, it is able to assert that this source is neither a monetary policy sector specified as a Taylor rule, nor a stylized private sector block with only inflation and real activity.

4.4. Sensitivity to priors and robustness to alternative specifications

This subsection shows robustness of the empirical results to alternative priors and specifications.

4.4.1. Priors and model's fit. The main results of Section 4 were presented for the particular prior specification chosen in Section 4.1. This subsection justifies this choice and demonstrates the robustness of the empirical conclusions to alternative prior specifications. While the choice of the priors for the initial states is completely innocuous,[19] the selection of k_Q, k_S and k_W turns out to be more important. Especially for k_Q this is not surprising: in a VAR with three variables and two lags the Q matrix contains 231 free parameters. With such a high number of free parameters, the specification of a sensible prior becomes essential, in order to prevent cases of ill-determinations like the ones described in Section 3.

While posterior inference can in principle be affected by the choice of k_Q, k_S and k_W, it is worth noting that k_Q, k_S and k_W do not parameterize time variation, but just prior beliefs about the amount of time variation. In large samples, as usual, the posterior mean converges to the maximum likelihood estimator. Consider as an example the matrix Q, which represents the

18. It is worth pointing out that the data reject also the hypothesis that changes in the conduct of monetary policy have influenced inflation and unemployment through an indirect channel, *i.e.* by affecting the variability of non-policy shocks. In fact, the correlation between the innovations to the reactions to inflation and unemployment and the innovations to the variances of inflation and unemployment are estimated to be essentially equal to zero (once the estimation algorithm is extended to take into account the possibility that these correlations are different from zero).

19. Much flatter specifications of these priors (for example, with variances 10 or 20 times bigger) deliver exactly the same results.

amount of time variation in the coefficients (B's). The random variable $Q \mid B^T$ is distributed as an inverse-Wishart. It can be shown that the conditional posterior mean has the following form:

$$E(Q \mid B^T) = \frac{v}{v+T}\frac{\overline{Q}}{v} + \frac{T}{v+T}Q^*, \tag{14}$$

where $\overline{Q}$ is the prior scale matrix, v are the prior degrees of freedom, Q^* is the maximum likelihood estimate of Q (conditional on B^T) and T is the sample size.[20] In this framework, the choice of k_Q would affect $\overline{Q}$. Given (14), it is easy to see that k_Q parameterizes time variation only when the prior degrees of freedom are driven to infinity. In general the posterior mean will be a combination of the prior and likelihood information, where the weights are determined by the relative size of the degrees of freedom of the prior and the sample size.

This being said, three factors lead to the choice of $k_Q = 0{\cdot}01$ in the benchmark model. First, the model seems to misbehave if a higher k_Q is chosen. It can be shown that, in the benchmark case of $k_Q = 0{\cdot}01$, the prior mean for Q implies a 95% probability of a 78% average cumulative change of the coefficients over the relevant sample. This is saying that $k_Q = 0{\cdot}01$ is a value that does not particularly penalize time variation in the coefficients. With a prior that explicitly favours a higher degree of time variation, the coefficients change considerably over time, but just in order to explain outliers and to push the in-sample error to zero. Their time variation captures much more high frequency variation than the one that would be captured by a pure random walk. In other words, most of the variation of the data seems to be explained by the raw shocks as opposed to the dynamics of the model.[21] A strange behaviour such as this is typical of very narrow likelihood peaks in very uninteresting regions of the parameter space, where the level of the likelihood is not informative of the model's fit. Moreover, any selection criterion based formally (Uhlig, 2001) or informally on the shape of the impulse responses would reject this model, which exhibits responses often exploding and with counterintuitive signs.

The second reason for choosing $k_Q = 0{\cdot}01$ is being consistent with the literature. Cogley and Sargent (2001) use the same value. Stock and Watson (1996) experiment with values of k_Q higher and lower than 0·01, pointing out that models with large *a priori* time variation do poorly in forecasting. Notice that using a lower k_Q would only reinforce the conclusion that changes in the conduct of monetary policy are not responsible for the economic instability of the 1970's.

The third reason that justifies the choice of $k_Q = 0{\cdot}01$ is a formal model selection. Proper posterior probabilities for a set of models (characterized by different prior assumptions on the time variation of the unobservable states) are computed. These are the probabilities of each model, once all the other parameters are integrated out. They are computed according to the reversible jump Markov chain Monte Carlo (RJMCMC) method described in Dellaportas, Forster and Ntzoufras (2002), that extends the original contribution of Carlin and Chib (1995). In the context of static and dynamic factor models, Lopes and West (2000) and Justiniano (2004) have shown that RJMCMC outperforms alternative methods, on the basis of the computation of the marginal likelihood (for an overview, see Justiniano, 2004). The details of the application to this model are provided in Appendix C. The posterior probabilities are computed for 18 models, represented by all possible combinations of $k_Q = \{0{\cdot}01; 0{\cdot}05; 0{\cdot}1\}$, $k_S = \{0{\cdot}01; 0{\cdot}1; 1\}$ and $k_W = \{0{\cdot}001; 0{\cdot}01\}$. Doing model selection to choose among models with different priors might seem unusual. However, notice that this procedure is equivalent to imposing a single, more complex prior, given by a mixture of the original, simple priors. Independently of the initial value

20. Remember that the scale matrix in the inverse-Wishart parameterization has the interpretation of sum of squared residuals. Therefore $\frac{\overline{Q}}{v}$ is interpretable as a variance and is comparable to Q^*.

21. The same type of misbehaviour appears if the prior degrees of freedom for Q are set to 22 (which is the minimum in order to have a proper prior), unless k_Q is also reduced to approximately 0·003. In this case the results are very similar to the ones of the benchmark case of Sections 4.2 and 4.3.

of the chain, the model characterized by $k_Q = 0{\cdot}01$, $k_S = 0{\cdot}1$ and $k_W = 0{\cdot}01$ gets a posterior probability that is essentially one. The fact that only one of the models gets positive posterior probability is not surprising and just indicates that the models' space is sparse. Observe that, while in some applications this is undesirable, here this is not a problem. In fact, the objective is just avoiding of imposing priors that are clearly at odds with the data. Finally, notice that the selected model is the one with the smallest value of k_Q. As argued above, this value of k_Q is not very restrictive against time variation in the coefficients of the lag structure. Yet, the results of Sections 4.2 and 4.3 seem to suggest that the amount of time variation of the coefficients of the lag structure is small relative to the time variation of the covariance matrix. Therefore, while in principle possible, it does not seem necessary to extend the model selection to include even smaller values of k_Q. A similar argument applies to high values of k_W.

Given that the introduction of a time varying A_t matrix is one of the contributions of this paper, there is no literature to refer to for a choice of k_S. Therefore, k_S is fixed to 0·1 (the value of the best fitting model) for the benchmark results. Also the choice of $k_W = 0{\cdot}01$ is motivated by the results of the model selection procedure. In addition, the model is estimated also for other possible values of k_S and k_W. The results (not reported) obtained with $k_S = 1$ are very similar to the benchmark case. Instead, when k_S is set to 0·01, the changes in the interest rate responses to inflation and real activity are smaller and smoother. If anything, this reinforces the conclusion that changes in the conduct of monetary policy are not responsible for the economic instability of the 1970's. If k_W is set to 0·001 the results are very much the same as in the baseline case. This supports the idea that time variation in the volatilities is fundamental to improving the fit of the model.

4.4.2. Random walk assumption. As discussed in Section 2, the random walk assumption for the evolution of the coefficients and log variances has some undesirable implications. In particular, it implies that the correlation among the reduced form residuals may become arbitrarily large in absolute value. To show that these undesirable implications are not to be taken too seriously, the model is re-estimated allowing the log standard errors and the coefficients of the simultaneous relations to follow more general AR processes. In a first experiment the autoregressive coefficients are all set to 0·95 (instead of 1 as in the random walk assumption), producing results not relevantly different from the ones presented above. In a second experiment these coefficients are estimated, by inserting an additional step in the Gibbs sampler.[22] Again, the results produced by such a model are in line with the results of the benchmark model. The AR coefficients for the log standard errors of the residuals are estimated to be very high (between 0·95 and 1), while the AR coefficients for the simultaneous relations are estimated to be lower (between 0·5 and 0·9). The only consequence of this fact is that the model captures many temporary parameter shifts, in addition to the permanent ones.

4.4.3. Cross-equation restrictions. The main results of Section 4 are presented for the case of S being block diagonal, with the block corresponding to parameters belonging to separate equations. This choice simplifies the inference, but may be seen as an important restriction. To verify that the results are not affected by such a choice, this assumption is now relaxed, allowing S to be unrestricted. This involves a modification of the basic Markov chain Monte Carlo algorithm, as reported in Appendix D. The model is re-estimated with this modified algorithm and the results are very similar to the main ones of Section 4.[23] The only notable difference is that in the case of

22. Drawing the states with the usual Gibbs procedure for state space forms, it is easy to draw the AR coefficients of the state evolution equation. The posterior is in fact normal, centred on the OLS coefficient with variance equal to the OLS variance.

23. The model is re-estimated using both the proposal distributions suggested in Appendix D. In both cases the acceptance rate of the Metropolis–Hastings step is approximately 12% and the estimates are very similar to each other.

an unrestricted S the estimates of the simultaneous relations exhibit slightly less time variation than in the benchmark case. Again, if anything, this strengthens the idea that changes in policy are not very important for explaining the outbursts of inflation and unemployment. Furthermore, the correlations of the innovations with the coefficients of different equations are estimated to be basically zero.

In order to get some further insight into the relevance of cross-equation restrictions for the application to the U.S. data, the model is re-estimated eliminating all the cross-equation restrictions, also in the lag structure. In practice, both S and Q are now assumed to be block diagonal, with the blocks corresponding to elements of the same equation. The results do not differ in any relevant way from the benchmark case, suggesting that cross-equation restrictions are not quantitatively very important.

5. CONCLUSIONS

This paper applies Markov chain Monte Carlo methods to the estimation of a time varying structural VAR. The sources of time variation are both the coefficients and, most importantly, the variance covariance matrix of the shocks. The time variation in the variance covariance matrix of the innovations not only seems to be empirically very important, but is also crucial for analysing the dynamics of the contemporaneous relations among the variables of the system.

In particular, the paper focuses on the role of monetary policy in the dynamics of inflation and unemployment for the U.S. economy. There is evidence of time variation in the U.S. monetary policy. Non-systematic policy has changed considerably over the last 40 years, becoming less important in the last part of the sample. Systematic monetary policy on the other hand has become more aggressive against inflation and unemployment. Nevertheless, there seems to be little evidence for a causal link between changes in interest rate systematic responses and the high inflation and unemployment episodes. It is important to stress that this is not a statement about neutrality of monetary policy. For example, it is quite conceivable that monetary policy which does not respond to inflation at all would introduce substantial instability in the economy. The estimates indicate, however, that within the range of policy parameters in the post-war era, the differences between regimes were not large enough to explain a substantial part of the fluctuations in inflation and unemployment. In particular, the less aggressive policy pursued by Burns in the 1970's was not loose enough to cause the great stagflation that followed. Indeed, the peaks in inflation and unemployment of the 1970's and early 1980's seem to be better explained by non-policy shocks than weaker interest rate responses to inflation and real activity.

In order to explore the true behavioural sources of these sequences of particularly bad shocks, a larger and fully structural model is needed. This suggests a clear direction for future research, namely allowing for a general form of simultaneity, overidentifying assumptions, cross-equation restrictions in the VAR representation and for a larger set of variables, solving the problem of the high number of parameters. A natural approach seems to be assuming the existence of common factors driving the dynamics of the coefficients.

APPENDIX A. THE BASIC MARKOV CHAIN MONTE CARLO ALGORITHM

Appendix A.1. Step 1: Drawing coefficient states

Conditional on A^T, Σ^T and V, the observation equation (4) is linear and has Gaussian innovations with known variance. As shown in Fruhwirth-Schnatter (1994) and Carter and Kohn (1994), the density $p(B^T \mid y^T, A^T, \Sigma^T, V)$ can be

The 12% acceptance rate seems acceptable, even though it is clear that the convergence is much slower than in the baseline case. This might also explain why the estimates of the simultaneous relations exhibit slightly less time variation than in the benchmark case.

factored as

$$p(B^T \mid y^T, A^T, \Sigma^T, V) = p(B_T \mid y^T, A^T, \Sigma^T, V) \prod_{t=1}^{T-1} p(B_t \mid B_{t+1}, y^t, A^T, \Sigma^T, V),$$

where

$$\begin{aligned} B_t \mid B_{t+1}, y^t, A^T, \Sigma^T, V &\sim N(B_{t|t+1}, P_{t|t+1}), \\ B_{t|t+1} &= E(B_t \mid B_{t+1}, y^t, A^T, \Sigma^T, V), \\ P_{t|t+1} &= \mathrm{Var}(B_t \mid B_{t+1}, y^t, A^T, \Sigma^T, V). \end{aligned}$$

$p(\cdot)$ is used to denote a generic density function, while N denotes the Gaussian distribution. The vector of B's can be easily drawn because $B_{t|t+1}$ and $P_{t|t+1}$ can be computed using the forward (Kalman filter) and the backward recursions reported in Appendix A.6, applied to the state space form given by (4) and (5). Specifically, the last recursion of the filter provides $B_{T|T}$ and $P_{T|T}$, *i.e.* the mean and variance of the posterior distribution of B_T. Drawing a value from this distribution, the draw is used in the backward recursion to obtain $B_{T-1|T}$ and $P_{T-1|T}$ and so on.[24]

Appendix A.2. Step 2: Drawing covariance states

The system of equations (4) can be written as

$$A_t(y_t - X_t' B_t) = A_t \widehat{y}_t = \Sigma_t \varepsilon_t, \tag{A.1}$$

where, taking B^T as given, $\widehat{y}_t$ is observable. Since A_t is a lower triangular matrix with ones on the main diagonal, (A.1) can be rewritten as

$$\widehat{y}_t = Z_t \alpha_t + \Sigma_t \varepsilon_t. \tag{A.2}$$

α_t is defined in (6) and Z_t is the following $n \times \frac{n(n-1)}{2}$ matrix:

$$Z_t = \begin{bmatrix} 0 & \cdots & \cdots & 0 \\ -\widehat{y}_{1,t} & 0 & \cdots & 0 \\ 0 & -\widehat{y}_{[1,2],t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\widehat{y}_{[1,\ldots,n-1],t} \end{bmatrix},$$

where, abusing notation, $\widehat{y}_{[1,\ldots,i],t}$ denotes the row vector $\left[\widehat{y}_{1,t}, \widehat{y}_{2,t}, \ldots, \widehat{y}_{i,t}\right]$.

Observe that the model given by (A.2) and (6) has a Gaussian but nonlinear state space representation. The problem, intuitively, is that the dependent variable of the observation equation, $\widehat{y}_t$, appears also on the R.H.S. in Z_t. Therefore, the vector $[\widehat{y}_t, \widehat{\alpha}_t]$ is not jointly normal and, thus, the conditional distributions cannot be computed using the standard Kalman filter recursion. However, under the additional assumption of S block diagonal, this problem can be solved by applying the Kalman filter and the backward recursion equation by equation. This is not only because the dependent variable of each equation, $\widehat{y}_{i,t}$, does not show up on the R.H.S. of the same equation, but because the vector $\widehat{y}_{[1,\ldots,i-1],t}$ can also be treated as predetermined in the i-th equation, due to the triangular structure. Another way to see this is to reinterpret time as going through equations, that is treating different measurement equations as belonging to distinct and subsequent time periods. Exactly like in the previous step of the sampler, this procedure allows one to recursively recover

$$\begin{aligned} \alpha_{i,t|t+1} &= E(\alpha_{i,t} \mid \alpha_{i,t+1}, y^t, B^T, \Sigma^T, V), \\ \Lambda_{i,t|t+1} &= \mathrm{Var}(\alpha_{i,t} \mid \alpha_{i,t+1}, y^t, B^T, \Sigma^T, V), \end{aligned}$$

where $\alpha_{i,t}$ is the i-th block of α_t, corresponding to the coefficients of the i-th equation in (A.2). As above, $\alpha_{i,t}$ can be drawn recursively from $p(\alpha_{i,t} \mid \alpha_{i,t+1}, y^t, B^T, \Sigma^T, V)$, which is $N(\alpha_{i,t|t+1}, \Lambda_{i,t|t+1})$.

Notice that without block diagonality of S it would not be possible to apply the recursions equation by equation and to separately draw every block α_i^T. However, while the S block diagonal assumption simplifies the analysis, it is not essential and Appendix D provides a modification of this algorithm in order to deal with an unrestricted S.

Appendix A.3. Step 3: Drawing volatility states

Consider now the system of equations

$$A_t(y_t - X_t' B_t) = y_t^* = \Sigma_t \varepsilon_t, \tag{A.3}$$

24. For further details on Gibbs sampling for state space models see Carter and Kohn (1994).

where, taking B^T and A^T as given, y_t^* is observable. This is a system of nonlinear measurement equations, but can be easily converted into a linear one, by squaring and taking logarithms of every element of (A.3). Due to the fact that $y_{i,t}^2$ can be very small, an offset constant is used to make the estimation procedure more robust. This leads to the following approximating state space form:

$$y_t^{**} = 2h_t + e_t \tag{A.4}$$

$$h_t = h_{t-1} + \eta_t. \tag{A.5}$$

$y_{i,t}^{**} = \log[(y_{i,t}^*)^2 + \bar{c}]$; $\bar{c}$ is the offset constant (set to 0·001); $e_{i,t} = \log(\varepsilon_{i,t}^2)$; $h_{i,t} = \log \sigma_{i,t}$. Observe that the e's and the η's are not correlated.[25]

The system in this form has a linear, but non-Gaussian state space form, because the innovations in the measurement equations are distributed as $\log \chi^2(1)$. In order to further transform the system into a Gaussian one, a mixture of normals approximation of the $\log \chi^2$ distribution is used, as described in Kim *et al.* (1998). Observe that the variance covariance matrix of the ε's is the identity matrix. This implies that the variance covariance matrix of the e's is also diagonal, allowing one to use the same (independent) mixture of normals approximation for any element of e. Kim *et al.* (1998) select a mixture of seven normal densities with component probabilities q_j, means $m_j - 1{\cdot}2704$, and variances v_j^2, $j = 1, \ldots, 7$. The constants $\{q_j, m_j, v_j^2\}$ are chosen to match a number of moments of the $\log \chi^2(1)$ distribution. Appendix A.7 reports the constants $\{q_j, m_j, v_j^2\}$.[26]

Define $s^T = [s_1, \ldots, s_T]'$, the matrix of indicator variables selecting at every point in time which member of the mixture of the normal approximation has to be used for each element of e. Conditional on B^T, A^T, V and s^T, the system has an approximate linear and Gaussian state space form. Again, exactly like in the previous steps of the sampler, this procedure allows one to recursively recover

$$h_{t|t+1} = E(h_t \mid h_{t+1}, y^t, A^T, B^T, V, s^T),$$
$$H_{t|t+1} = \mathrm{Var}(h_t \mid h_{t+1}, y^t, A^T, B^T, V, s^T)$$

and recursively draw every h_t from $p(h_t \mid h_{t+1}, y^t, A^T, B^T, V, s^T)$, which is $N(h_{t|t+1}, H_{t|t+1})$.[27]

Conditional on y^{**T} and the new h^T, it is possible to sample the new s^T matrix, to be used in the next iteration. As in Kim *et al.* (1998), this is easily done by independently sampling each $s_{i,t}$ from the discrete density defined by

$$\Pr(s_{i,t} = j \mid y_{i,t}^{**}, h_{i,t}) \propto q_j \, f_N(y_{i,t}^{**} \mid 2h_{i,t} + m_j - 1{\cdot}2704, v_j^2), \qquad j = 1, \ldots, 7, \; i = 1, \ldots, n.$$

Appendix A.4. Step 4: Drawing hyperparameters

The hyperparameters of the model are the diagonal blocks of V: Q, W and the diagonal blocks of S (corresponding to parameters belonging to different equations). Conditional on B^T, Σ^T, A^T and y^T, each square block has an inverse-Wishart posterior distribution, independent of the other blocks. Moreover, conditional on B^T, Σ^T, A^T and y^T, it is easy to draw from these inverse-Wishart posteriors because the innovations are observable.[28]

Appendix A.5. Summary

To summarize, the sampler takes the form:

a. Initialize A^T, Σ^T, s^T and V.
b. Sample B^T from $p(B^T \mid y^T, A^T, \Sigma^T, V)$.

25. This is easy to see since the ε's and the η's are independent. However, even if the ε's and the η's were not independent, the e's and the η's would be uncorrelated. For a proof of this result, see Harvey *et al.* (1994).

26. In this paper the reweighting procedure used in Kim *et al.* (1998) to correct the minor approximation error is not adopted.

27. Cogley (2003) and Cogley and Sargent (2003) have a similar model for the variances. However, to evaluate the posterior they adopt the algorithm suggested by Jacquier, Polson and Rossi (1994), that draws the volatility states one at a time. As pointed out in Carter and Kohn (1994) and Kim *et al.* (1998), this drastically increases the autocorrelations of the draws and decreases efficiency. Uhlig (1997) has a similar model with constant coefficients. However, he adopts nonlinear filtering procedures and importance sampling to evaluate the posterior. As stressed by Uhlig himself, importance sampling procedures are subject to the risk of real inefficient sampling, in the cases in which, out of thousands of draws, only a few of them have a non-zero weight.

28. See Gelman, Carlin, Stern and Rubin (1995) for a description of the sampling procedure from a normal-inverse-Wishart distribution, also when combined with natural conjugate priors.

c. Sample A^T from $p(A^T \mid y^T, B^T, \Sigma^T, V)$.

d. Sample Σ^T from $p(\Sigma^T \mid y^T, A^T, B^T, s^T, V)$.

e. Sample s^T from $p(s^T \mid y^T, A^T, \Sigma^T, V)$.

f. Sample V, by sampling Q, W and S from $p(Q, W, S \mid y^T, B^T, A^T, \Sigma^T) = p(Q \mid y^T, B^T, A^T, \Sigma^T) \cdot p(W \mid y^T, B^T, A^T, \Sigma^T) \cdot p(S_1 \mid y^T, B^T, A^T, \Sigma^T) \cdot \ldots \cdot p(S_{n-1} \mid y^T, B^T, A^T, \Sigma^T)$.

g. Go to b.

Appendix A.6. Gibbs sampling for state space models

Consider a measurement equation:

$$y_t = H_t \beta_t + \varepsilon_t$$

and a transition equation

$$\beta_t = F\beta_{t-1} + u_t$$

where

$$\begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix} \sim \text{i.i.d. } N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} R_t & 0 \\ 0 & Q \end{bmatrix}\right).$$

Let

$$\begin{aligned} \beta_{t|s} &= E(\beta_t \mid Y^s, H^s, R^s, Q), \\ V_{t|s} &= \text{Var}(\beta_t \mid Y^s, H^s, R^s, Q). \end{aligned}$$

Then, given $\beta_{0|0}$ and $V_{0|0}$, a standard Kalman filter delivers:

$$\begin{aligned} \beta_{t|t-1} &= F\beta_{t-1|t-1}, \\ V_{t|t-1} &= FV_{t-1|t-1}F' + Q, \\ K_t &= V_{t|t-1}H_t'(H_tV_{t|t-1}H_t' + R_t)^{-1}, \\ \beta_{t|t} &= \beta_{t|t-1} + K_t(y_t - H_t\beta_{t|t-1}), \\ V_{t|t} &= V_{t|t-1} - K_tH_tV_{t|t-1}. \end{aligned}$$

The last elements of the recursion are $\beta_{T|T}$ and $V_{T|T}$, which are the mean and the variance of the normal distribution used to make a draw for β_T. The draw of β_T and the output of the filter are now used for the first step of the backward recursion, which provides $\beta_{T-1|T}$ and $V_{T-1|T}$, used to make a draw of β_{T-1}. The backward recursion continues until time zero. For a generic time t, the updating formulas of the backward recursion are:

$$\begin{aligned} \beta_{t|t+1} &= \beta_{t|t} + V_{t|t}F'V_{t+1|t}^{-1}(\beta_{t+1} - F\beta_{t|t}), \\ V_{t|t+1} &= V_{t|t} - V_{t|t}F'V_{t+1|t}^{-1}FV_{t|t}. \end{aligned}$$

Appendix A.7. Selection of the mixing distribution to be $\log \chi^2(1)$*:*

ω	$q_j = \Pr(\omega = j)$	m_j	v_j^2
1	0·00730	−10·12999	5·79596
2	0·10556	−3·97281	2·61369
3	0·00002	−8·56686	5·17950
4	0·04395	2·77786	0·16735
5	0·34001	0·61942	0·64009
6	0·24566	1·79518	0·34023
7	0·25750	−1·08819	1·26261

Source: Kim *et al*. (1998).

APPENDIX B. CONVERGENCE OF THE MARKOV CHAIN MONTE CARLO ALGORITHM

This appendix assesses convergence of the Markov chain Monte Carlo algorithm in the baseline application to the U.S. data.[29]

First of all, different starting points of the chain (randomly selected) produce the same results. These results are also not sensitive to the initial number of discarded draws or the total number of passes of the Gibbs sampler (an experiment with 30,000 iterations, keeping one every three draws, gave the same results). For space reasons, a complete description of the characteristics of the chain is provided only for 706 parameters of the model: the hyperparameters (the 241 free elements of V) and the volatilities (three sets of volatility states for 155 time periods).

In order to judge how well the chain mixes, common practice is to look at the autocorrelation function of the draws. Low autocorrelations suggest that the draws are almost independent, which increases the efficiency of the algorithm. Figure 9(a) plots the 20-th-order sample autocorrelation of the draws. In all three panels of Figure 9, points 1–241 correspond to the free elements of V (first the variances and then the covariances by row), while points 242–706 correspond to the volatilities (for every time period, by equation). With some exceptions the plot remains below 0·2 and often below 0·1. Closely related, Figure 9(b) presents the inefficiency factors (IFs) for the posterior estimates of the parameters. The IF is the inverse of the relative numerical efficiency measure of Geweke (1992), *i.e.* the IF is an estimate of $\left(1 + 2\sum_{k=1}^{\infty} \rho_k\right)$, where ρ_k is the k-th autocorrelation of the chain. In this application the estimate is performed using a 4% tapered window for the estimation of the spectral density at frequency zero. Values of the IFs below or around 20 are regarded as satisfactory. Again, except for a few cases, the IFs are around or even below 20. The table below summarizes the distribution of the IFs for the posterior estimates of all different sets of parameters. On average, the set of parameters with the highest IFs are the A's, for which it is not uncommon to have IFs between 4 and 75. In all other cases, 90% of the IFs are below 25. The IFs are remarkably low for the time varying coefficients (B's).

	Median	Mean	Min	Max	10-th Percentile	90-th Percentile
V	18·99	22·29	9·63	185·51	12·76	25·60
Σ	7·49	9·61	1·62	45·61	4·31	18·80
A	20·01	30·32	2·07	118·67	4·85	75·81
B	5·63	6·86	2·34	18·50	2·94	13·61

Summary of the distribution of the inefficiency factors for different sets of parameters. V: elements of the covariance matrix of the model's innovations; Σ: time varying volatilities; A: time varying simultaneous relations; B: time varying coefficients.

As a further check, Figure 9(c) plots the Raftery and Lewis (1992) diagnostic of the total number of runs required to achieve a certain precision (the parameters for the diagnostic are specified as follows: quantile = 0·025; desired accuracy = 0·025; required probability of attaining the required accuracy = 0·95). The required number of runs is always far below the total number of iterations of this application. To conclude, considering the very high dimensionality of the problem, the convergence diagnostics seem satisfactory.

APPENDIX C. REVERSIBLE JUMP MARKOV CHAIN MONTE CARLO (RJMCMC)

This appendix describes the RJMCMC algorithm used for the model selection. Further details on the method are in Lopes and West (2000), Dellaportas *et al*. (2002) and Justiniano (2004).

Consider a set M of competing models. In the case of this paper different models are characterized by different prior assumptions on the time variation of the unobservable states, as described in Section 4.1. For every model $m \in M$, define the set of model's parameters

$$\theta^m = \{\alpha_1^m, \ldots, \alpha_T^m, \log \sigma_1^m, \ldots, \log \sigma_T^m, Q^m, S^m, W^m\}.$$

In very general terms, the method consists of simulating the joint posterior probability distribution of the couple (m, θ^m). This can be done using a standard Metropolis–Hastings approach. A proposal value $\left(m', \theta^{m'}\right)$ is generated from a proposal distribution $q_p\left(m', \theta^{m'}\right) = q\left(\theta^{m'} \mid m'\right) \cdot J(m')$. The new value is accepted with the usual Metropolis–Hastings acceptance probability (stated below). If the proposal value is rejected, it is replaced by the previous element of the chain. This procedure generates a set of values, which can be used to approximate the posterior of (m, θ^m) and, among other things, to compute posterior probabilities over the models' space.

29. For the convergence checks, I used modified versions of codes of Alejandro Justiniano and James P. LeSage.

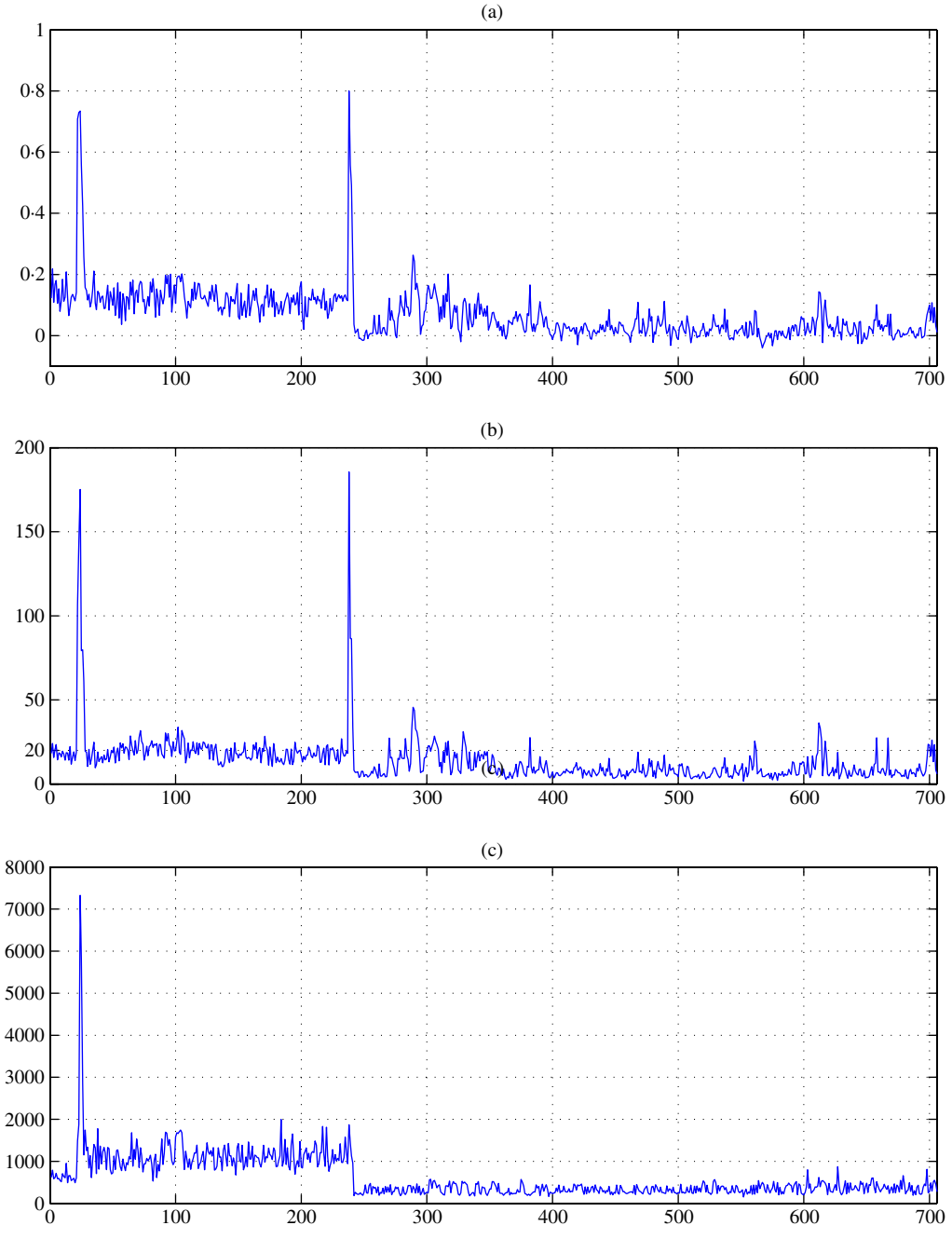

FIGURE 9

Convergence diagnostics for hyperparameters and volatilities. (a) 20-th-order sample autocorrelation, (b) inefficiency factors, (c) Raftery and Lewis's total number of runs. See Appendix B for details

The details of the algorithm adopted in the paper are as follows:

a. For every model $m \in M$, the output of a preliminary run of the Gibbs sampler is used to construct the following,

independent proposal distributions for the elements of θ^m:

$$\begin{aligned} q\left(\alpha_t^m \mid m\right) &= N\left(\overline{\alpha_t^m}, 2 \cdot \overline{\text{var}(\alpha_t^m)}\right), \\ q\left(\log \sigma_t^m \mid m\right) &= N\left(\overline{\log \sigma_t^m}, 2 \cdot \overline{\text{var}(\log \sigma_t^m)}\right), \\ q\left(Q^m \mid m\right) &= IW\left(40 \cdot \overline{Q^m}, 40\right), \\ q\left(W^m \mid m\right) &= IW\left(4 \cdot \overline{W^m}, 4\right), \\ q\left(S_1^m \mid m\right) &= IW\left(2 \cdot \overline{S_1^m}, 2\right), \\ &\vdots \\ q\left(S_{n-1}^m \mid m\right) &= IW\left((n-1) \cdot \overline{S_{n-1}^m}, n-1\right). \end{aligned}$$

The variables denoted by the upper bar are posterior means of the preliminary Gibbs sampler, except for $\overline{\text{var}(\alpha_t^m)}$ and $\overline{\text{var}(\log \sigma_t^m)}$ which are posterior variances. In practice, the preliminary Gibbs run is used to centre approximate posteriors, that will be adopted as proposal distributions. The approximate posteriors are made more diffuse than the exact ones on purpose, in order to facilitate the convergence of the algorithm. In this paper the preliminary Gibbs samplers are based on 2000 iterations, discarding the first 500.

b. Initialize m and draw a value of θ^m from the proposal distribution $q\left(\theta^m \mid m\right)$.

c. Draw a candidate couple $\left(m', \theta^{m'}\right)$ from $q_p\left(m', \theta^{m'}\right)$. This can be easily done by first drawing m' from a proposal distribution $J(m')$ over the models' space and then $\theta^{m'}$ from $q\left(\theta^{m'} \mid m'\right)$.

d. Accept the new couple $\left(m', \theta^{m'}\right)$ with probability

$$a = \min\left\{\frac{p(y^T \mid m', \theta^{m'})p(\theta^{m'} \mid m')p(m')}{p(y^T \mid m, \theta^m)p(\theta^m \mid m)p(m)} \frac{q(\theta^m \mid m)J(m)}{q(\theta^{m'} \mid m')J(m')}, 1\right\},$$

where $p(y^T \mid m, \theta^m)$ is the likelihood, computed through the Kalman filter to integrate out the B's; $p(\theta^m \mid m)$ is the prior of θ^m within model m; $p(m)$ is the prior probability of model m; $q(\theta^m \mid m)$ is the proposal distribution for the parameters, conditional on the model, constructed as above; $J(m)$ is the unconditional proposal distribution over the models. If the new draw is not accepted, keep the previous couple $\left(m, \theta^m\right)$.

e. Go to c.

APPENDIX D. THE MARKOV CHAIN MONTE CARLO ALGORITHM IN THE CASE OF S UNRESTRICTED

This appendix presents an extension of the basic MCMC algorithm, produced in order to deal with the case of S (the variance covariance matrix of the innovations to the simultaneous relations) left unrestricted. The modification works as follows: step 2 of the Gibbs sampler is replaced by a Metropolis–Hastings step in which a candidate value for the α's, α^{T*}, is drawn from a proposal distribution $f_P(\alpha^T \mid y^T, B^T, \Sigma^T, V)$. The candidate value is accepted with probability equal to

$$\min\left\{1, \frac{p(\alpha^{T*} \mid y^T, B^T, \Sigma^T, V) f_P(\alpha^{T,-1} \mid y^T, B^T, \Sigma^T, V)}{p(\alpha^{T,-1} \mid y^T, B^T, \Sigma^T, V) f_P(\alpha^{T*} \mid y^T, B^T, \Sigma^T, V)}\right\},$$

where $p(\cdot)$ is the posterior of α^T and $\alpha^{T,-1}$ represents the previous draw of α^T in the chain. If the candidate is rejected, α^T is set to be $\alpha^{T,-1}$. Step 4 is modified straightforwardly in order to sample S jointly and not in blocks. As in the Gibbs sampling case, this procedure allows one to draw from the exact joint posterior of the parameters. The last issue is related to the choice of the proposal distribution. A natural candidate is the distribution used to sample α^T under the assumption of S block diagonal. Another possibility is to treat (erroneously) the state space form given by (A.2) and (6) as linear and draw a candidate value for α^T using the usual forward and backward recursions. Clearly, if the proposal distribution is not close to the true density, convergence can be much slower than in the baseline case. In the context of the empirical application of this paper, with both the proposal distributions suggested above, the acceptance rate was approximately 12%.

Acknowledgements. This paper is a revised version of the first chapter of my Ph.D. dissertation produced at Princeton University. I am indebted to Chris Sims, whose suggestions considerably improved the quality of the paper.

I am also grateful to Timothy Cogley, Giovanni Favara, Paolo Giordani, Hidehiko Ichimura, Alejandro Justiniano, Siem Jan Koopman, Michele Lenza, Guido Lorenzoni, Bernard Salanié, Paolo Surico, Lars Svensson, Andrea Tambalotti, Thijs van Rens, especially Mark Watson and two anonymous referees for many helpful comments and discussions. Remaining errors are my own.

REFERENCES

BERNANKE, B. S. and MIHOV, I. (1998), "Measuring Monetary Policy", *The Quarterly Journal of Economics*, **113**, 869–902.
BLANCHARD, O. and SIMON, J. (2001), "The Long and Large Decline in U.S. Output Volatility", in *Brookings Papers on Economic Activity 1* 135–164.
BOIVIN, J. (2001), "The Fed's Conduct of Monetary Policy: Has It Changed and Does It Matter?" (Mimeo, Columbia Business School).
BOIVIN, J. and GIANNONI, M. (2003), "Has Monetary Policy Become More Effective?" (NBER Working Paper No. 9459).
BULLARD, J. and MITRA, K. (2000), "Determinacy, Learnability, and Monetary Policy Inertia" (Working Paper No. 200-030A, Federal Reserve Bank of St. Louis).
CANOVA, F. (1993), "Modelling and Forecasting Exchange Rates with a Bayesian Time-Varying Coefficient Model", *Journal of Economic Dynamics and Control*, **17**, 233–261.
CARLIN, B. P. and CHIB, S. (1995), "Bayesian Model Choice via Markov Chain Monte Carlo Methods", *Journal of Royal Statistical Society B*, **157**, 473–484.
CARTER, C. K. and KOHN, R. (1994), "On Gibbs Sampling for State Space Models", *Biometrika*, **81**, 541–553.
CHIB, S., NARDARI, F. and SHEPHARD, N. (2002), "Analysis of High Dimensional Multivariate Stochastic Volatility Models" (Mimeo, Nuffield College, University of Oxford).
CHRISTIANO, L. J., EICHENBAUM, M. and EVANS, C. (1999), "Monetary Shocks: What Have We Learned and To What End?", in J. B. Taylor and M. Woodford (eds.) *Handbook of Macroeconomics*, Vol. 1A (New York: Elsevier Science, North-Holland) 65–148.
CICCARELLI, M. and REBUCCI, A. (2003), "Measuring Contagion with a Bayesian Time-Varying Coefficient Model" (Mimeo, International Monetary Fund).
CLARIDA, R., GALÍ, J. and GERTLER, M. (2000), "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory", *The Quarterly Journal of Economics*, **115** (1), 147–180.
COGLEY, T. (2003), "How Fast Can the New Economy Grow? A Bayesian Analysis of the Evolution of Trend Growth" (Mimeo, University of California, Davis).
COGLEY, T. and SARGENT, T. J. (2001), "Evolving Post-World War II U.S. Inflation Dynamics", in *NBER Macroeconomics Annual* (Cambridge, MA: MIT Press) 331–373.
COGLEY, T. and SARGENT, T. J. (2003), "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S." (Mimeo, New York University).
COX, D. R. and WERMUTH, N. (1996) *Multivariate Dependencies—Models, Analysis and Interpretation* (London: Chapman and Hall).
DELLAPORTAS, P., FORSTER, J. J. and NTZOUFRAS, I. (2002), "On Bayesian Model and Variable Selection Using MCMC", *Statistics and Computing*, **12**, 27–36.
DEMPSTER, A. (1972), "Covariance Selection", *Biometrics*, **28**, 157–175.
DOUCET, A., DE FREITAS, N. and GORDON, N. (2001) *Sequential Monte Carlo Methods in Practice* (New York: Springer-Verlag).
FAVERO, C. A. and ROVELLI, R. (2003), "Macroeconomic Stability and the Preferences of the Fed. A Formal Analysis 1961–1998", *Journal of Money, Credit and Banking*, **35** (4), 545–556.
FRUHWIRTH-SCHNATTER, S. (1994), "Data Augmentation and Dynamic Linear Models", *Journal of Time Series Analysis*, **15**, 183–202.
GELMAN, A., CARLIN, J. B., STERN, H. S. and RUBIN, D. B. (1995) *Bayesian Data Analysis* (London: Chapman and Hall).
GEWEKE, J. (1992), "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments", in J. M. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith (eds.) *Bayesian Statistics* (Oxford: Oxford University Press) 169–193.
HAMILTON, J. H. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, **57** (2), 357–384.
HANSON, M. (2003), "Varying Monetary Policy Regimes: A Vector Autoregressive Investigation" (Mimeo, Wesleyan University).
HARVEY, A. C., RUIZ, E. and SHEPHARD, N. (1994), "Multivariate Stochastic Variance Models", *Review of Economic Studies*, **61**, 247–264.
JACQUIER, E., POLSON, N. G. and ROSSI, P. E. (1994), "Bayesian Analysis of Stochastic Volatility Models", *Journal of Business and Economic Statistics*, **12**, 371–417.
JACQUIER, E., POLSON, N. G. and ROSSI, P. E. (1995), "Models and Priors for Multivariate Stochastic Volatility" (CIRANO Working Paper 1995-18).
JUDD, J. and RUDEBUSCH, G. D. (1998), "Taylor's Rule and the Fed: 1970–1997", in *Economic Review*, Vol. 3 (Federal Reserve Bank of San Francisco) 3–16.

JUSTINIANO, A. (2004), "Estimation and Model Selection in Dynamic Factor Analysis" (Mimeo, International Monetary Fund).
KIM, C. and NELSON, C. R. (1999) *State-Space Models with Regime Switching* (Cambridge, MA: MIT Press).
KIM, S., SHEPHARD, N. and CHIB, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models", *Review of Economic Studies*, **65**, 361–393.
LEEPER, E., SIMS, C. A. and ZHA, T. (1996), "What Does Monetary Policy Do?", in *Brookings Papers on Economic Activity 2*, 1–78.
LEEPER, E. and ZHA, T. (2002), "Modest Policy Intervention" (NBER Working Paper No. 9192).
LOPES, H. F. and WEST, M. (2000), "Model Uncertainty in Factor Analysis" (Mimeo, Duke University).
LUBIK, T. A. and SCHORFHEIDE, F. (2004), "Testing for Indeterminacy: An Application to U.S. Monetary Policy", *American Economic Review*, **94** (1), 190–217.
LUCAS, R. E. (1976), "Econometric Policy Evaluation: A Critique", in K. Brunner and A. Meltzer (eds.) *The Phillips Curve and Labor Markets*, Vol. 1 (Carnegie-Rochester Series on Public Policy) 19–46.
PINHEIRO, J. and BATES, D. (1996), "Unconstrained Parametrizations for Variance–Covariance Matrices", *Statistics and Computing*, **11**, 289–296.
POURAHMADI, M. (1999), "Joint Mean–Covariance Models with Applications to Longitudinal Data: Unconstrained Parameterizations", *Biometrika*, **86**, 677–690.
POURAHMADI, M. (2000), "Maximum Likelihood Estimation for Generalized Linear Models for Multivariate Normal Covariance Matrices", *Biometrika*, **87**, 425–435.
RAFTERY, A. E. and LEWIS, S. (1992), "How Many Iterations in the Gibbs Sampler?", in J. M. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith (eds.) *Bayesian Statistics* (Oxford: Oxford University Press) 763–773.
ROTEMBERG, J. J. and WOODFORD, M. (1997), "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy", in *NBER Macroeconomics Annual* (Cambridge, MA: MIT Press) 297–346.
SARGAN, J. D. and BHARGAVA, A. (1983), "Maximum Likelihood Estimation of Regression Models with First Order Moving Average Errors when the Root Lies on the Unit Circle", *Econometrica*, **51**, 799–820.
SHEPHARD, N. (1994*a*), "Local Scale Models: State Space Alternatives to Integrated GARCH Processes", *Journal of Econometrics*, **60**, 181–202.
SHEPHARD, N. (1994*b*), "Partial Non-Gaussian State Space", *Biometrika*, **81**, 115–131.
SHEPHARD, N. (1996), "Statistical Aspects of ARCH and Stochastic Volatility", in D. R. Cox, D. V. Hinkley and O. E. Barndorff-Nielsen (eds.) *Time Series Models in Econometrics, Finance and Other Fields* (London: Chapman and Hall) 1–67.
SHEPHARD, N. and HARVEY, A. C. (1990), "On the Probability of Estimating a Deterministic Component in the Local Level Model", *Journal of Time Series Analysis*, **11** (4), 339–347.
SIMS, C. A. (1993), "A 9 Variable Probabilistic Macroeconomic Forecasting Model", in J. H. Stock and M. W. Watson (eds.) *Business Cycles, Indicators and Forecasting, NBER Studies in Business Cycles*, Vol. 28, 179–214.
SIMS, C. A. (1999), "Drift and Breaks in Monetary Policy" (Mimeo, Princeton University).
SIMS, C. A. (2000), "Using a Likelihood Perspective to Sharpen Econometric Discourse: Three Examples", *Journal of Econometrics*, **95** (2), 443–462.
SIMS, C. A. (2001*a*), "Stability and Instability in US Monetary Policy Behavior" (Mimeo, Princeton University).
SIMS, C. A. (2001*b*), "Comment on Sargent and Cogley's: 'Evolving Post World War II U.S. Inflation Dynamics' ", in *NBER Macroeconomics Annual* (Cambridge, MA: MIT Press) 373–379.
SIMS, C. A. and ZHA, T. (1998), "Bayesian Methods for Dynamics Multivariate Models", *International Economic Review*, **39**, 949–968.
SIMS, C. A. and ZHA, T. (2004), "Were There Regime Switches in US Monetary Policy?" (Mimeo, Princeton University).
SMITH, M. and KOHN, R. (2002), "Parsimonious Covariance Matrix Estimation for Longitudinal Data", *Journal of the American Statistical Association*, **97**, 1141–1153.
STOCK, J. H. and WATSON, M. W. (1996), "Evidence on Structural Instability in Macroeconomic Time Series Relations", *Journal of Business and Economic Statistics*, **14**, 11–30.
STOCK, J. H. and WATSON, M. W. (1998), "Asymptotically Median Unbiased Estimation of Coefficient Variance in a Time Varying Parameter Model", *Journal of the American Statistical Association*, **93**, 349–358.
STOCK, J. H. and WATSON, M. W. (2001), "Vector Autoregressions", *Journal of Economic Perspectives*, **15**, 101–116.
STOCK, J. H. and WATSON, M. W. (2002), "Has the Business Cycle Changed and Why?" (NBER Working Paper No. 9127).
UHLIG, H. (1997), "Bayesian Vector Autoregressions with Stochastic Volatility", *Econometrica*, **65**, 59–73.
UHLIG, H. (2001), "What Are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure" (Mimeo, Tilburg University).
WOODFORD, M. (2001), "The Taylor Rule and Optimal Monetary Policy", *American Economic Review*, **91** (2), 232–237.
WOODFORD, M. (2003) *Interest and Prices: Foundations of a Theory of Monetary Policy* (Princeton, NJ: Princeton University Press).