THE RIMINI CENTRE FOR ECONOMIC ANALYSIS

WP 24_08

Gary Koop
University of Strathclyde, UK
and
The Rimini Centre for Economic Analysis

Roberto Leon-Gonzalez
National Graduate Institute for Policy Studies, UK
and
The Rimini Centre for Economic Analysis

Rodney W. Strachan
University of Queensland, Australia
and
The Rimini Centre for Economic Analysis

"ON THE EVOLUTION OF MONETARY POLICY"

On the Evolution of Monetary Policy

Gary Koop[*]
University of Strathclyde

Roberto Leon-Gonzalez
National Graduate Institute for Policy Studies

Rodney W. Strachan
University of Queensland

August 18, 2007

_____

[*]Corresponding author: Department of Economics, University of Strathclyde, 130 Rottenrow, Glasgow G4 0GE, UK. Email: Gary.Koop@strath.ac.uk. All authors are Fellows of the Rimini Centre for Economic Analysis.

1

ABSTRACT

This paper investigates the evolution of monetary policy in the U.S. using a standard set of macroeconomic variables. Many recent papers have addressed the issue of whether the monetary transmission mechanism has changed (e.g. due to the Fed taking a more aggressive stance against inflation) or whether apparent changes are simply due to changes in the volatility of exogenous shocks. A subsidiary question is whether any such changes have been gradual or abrupt. In this paper, we shed light on these issues using a mixture innovation model which extends the class of time varying Vector Autoregressive models with stochastic volatility which have been used in the past. The advantage of our extension is that it allows us to estimate whether, where, when and how parameter change is occurring (as opposed to assuming a particular form of parameter change). Our empirical results strongly indicate that the transmission mechanism, the volatility of exogenous shocks and the correlations between exogenous shocks are all changing (albeit at different times and to different extents) Furthermore, evolution of parameters is gradual.

Keywords: structural VAR, monetary policy, Bayesian, mixture innovation model, time varying parameter model

JEL Classification: C11, C32, E52

1 Introduction

Questions of interest to policymakers typically involve the inter-relationships between several macroeconomic variables. To investigate such questions, it is common to build a macroeconomic model (e.g. based on a Vector Autoregressive, VAR, model) where exogenous shocks impact on the variables under study. The manner in which the exogenous variables affect the variables of interest is referred to as the transmission mechanism. Traditionally, estimation of the transmission mechanism (or features such as impulse responses which shed light on it) was considered a major goal of many macroeconomic papers. However, empirical researchers have realized two important things. First, the transmission mechanism may not be constant over time. Second, the way the exogenous shocks are generated (and, in particular, their variance) can change over time.

Consider, for instance, U.S. monetary policy and the question of whether the macroeconomic events of the 1970s were due to bad policy or bad luck. Some authors (e.g. Boivin and Giannoni, 2006, Cogley and Sargent, 2001 and Lubik and Schorfheide, 2004) have argued that the way the Fed reacted to inflation has changed over time (e.g. under the Volcker and Greenspan chairmanship, the Fed was more aggressive in fighting inflation pressures than under Burns). This is the "bad policy" story and is an example of a change in the transmission mechanism. Others (e.g. Sims and Zha, 2006) have emphasized that the variance of the exogenous shocks has changed over time and that this alone may explain many apparent changes in monetary policy. This is the "bad luck" story. Yet others (e.g. Primiceri, 2005) have found that both the transmission mechanism and the variance of the exogenous shocks has changed over time.

This brief (and very incomplete) discussion of the literature is intended to motivate the basic point that an understanding of monetary policy should be based on multivariate models where the transmission mechanism and the variances of the exogenous shocks can both potentially change over time. Another important issue is whether any such change is gradual or abrupt. Many models have been used to investigate such issues in the literature. However, most of them (including some of the DSGE-based models), use extended versions of VARs as building blocks. There is a large literature (e.g. George, Sun and Ni, 2006) which points out that even standard VARs can be over-parameterized and tries to find various ways of minimizing this problem. When one turns to extensions of VARs with time-varying parameters such

over-parameterization worries become even more serious. Such considerations motivate the present paper. In it, we re-examine some of the existing empirical literature on U.S. monetary policy using a class of models which is flexible enough to nest many of the existing specifications, but is more tightly parameterized in key dimensions. Most importantly, it allows us to estimate the form and nature of how parameters (and, thus, the transmission mechanism) evolve over time.

Our model is based on a time-varying VAR similar to that used in Primiceri (2005) or Cogley and Sargent (2001, 2005), but extends this type of model in important ways. Like Primiceri (2005) and Cogley and Sargent (2005), we have a multivariate model where both the transmission mechanism and the error covariance matrix can change over time. However, unlike Primiceri (2005) and the related time-varying parameter VAR (TVP-VAR) literature (e.g. Cogley and Sargent, 2001, 2005 and Cogley, Morozov and Sargent, 2005), we do not impose as many restrictions on the time variation of the parameters. Instead, to model the change in parameters over time, we draw on the mixture innovation approach of Gerlach, Carter and Kohn (2000) and Giordani and Kohn (2006) as a way of letting the data speak about how parameters evolve as well as keeping the model more tightly parameterized in key dimensions. Exact details will be provided in the next section. But, to motivate the basic ideas, note that there are two main approaches to modelling changes in parameters over time: one can estimate a model with a small number of structural breaks (usually one or two). Alternatively, one can estimate a time varying parameter (TVP) model where the parameters are allowed to change with each new observation, usually according to a random walk. A TVP model can be interpreted as imposing $T - 1$ breaks in a sample of size T. Thus, we have two extremes: models with very few (but usually large) breaks or those with many (usually small) breaks. The approach adopted in this paper allows for the estimation of the number of breaks. Thus, we nest the two extreme cases and can let the data tell us if there are few (or no) changes in the parameters or whether change is constant and gradual. Another advantage of our approach relative to the TVP-VAR literature is that, by estimating the parameters to be constant over periods, we can obtain a more parsimonious model, mitigating concerns about over-parameterization. Our model also allows for the three different blocks of parameters we work with (the VAR coefficients, a block which relates to the error variances and another relating to error covariances) to evolve in completely different ways (or even for some or all blocks not to change at

4

all). Thus, we can estimate whether and how change occurs in a very flexible manner, as opposed to assuming a specific model with parameter change of a particular sort.

After developing our model and appropriate Bayesian econometric methods, we present empirical results. We work with a standard system involving inflation, unemployment and interest rates. We present results relating to the transmission mechanism and the volatility of exogenous shocks. We find evidence of gradual change in all of our parameters and reinforce the findings of Primiceri (2005). Relative to the existing literature, a crude summary of our results might run as follows. The model of Primiceri (2005) is best, but the model of Cogley and Sargent (2005) is not too bad (although there are some restrictions in this model which are rejected, these have only minor macroeconomic implications). Models which only have time variation in the error covariance matrix (i.e. with constant VAR coefficients) are a bit worse. They accurately recover patterns in the exogenous shocks, but can be misleading about the transmission mechanism. However, models with a constant error covariance matrix such as Cogley and Sargent (2001) or a traditional VAR are strongly rejected and can yield seriously misleading policy inferences.

2   The Models

The models used in this paper all begin with a state space model involving a measurement equation:

$$y_t = Z_t \alpha_t + \varepsilon_t \tag{1}$$

and a state equation

$$\alpha_{t+1} = \alpha_t + R_t \eta_t, \tag{2}$$

where y_t is an $p \times 1$ vector of observations on the dependent variables, α_t an $m \times 1$ vector of states (in our case, these are the VAR coefficients), ε_t are independent $N(0, H_t)$ random vectors and η_t are independent $N(0, Q_t)$ random vectors for $t = 1, .., T$. The errors in the two equations, ε_t and η_s, are independent of one another for all t and s.[1] Z_t is the appropriate $p \times m$ matrix of data on explanatory variables. In our case, we are working with extensions

---

[1]This is the standard assumption, but it can easily be relaxed if desired.

of VARs and, hence, each row of Z_t contains lags of all dependent variables and an intercept and other deterministic terms. For future reference, note that we will use R_t to control the structural breaks in our model.

This model, which is a familiar one in the state space literature, nests a wide range of commonly-used models. A VAR is obtained if we set $R_t = 0_m$ for all t and, thus, the VAR coefficients are constant over time. TVP-VARs of the sort used, e.g., in Cogley and Sargent (2001) are obtained by setting $R_t = I_m$ for all t. Technical details on how Bayesian econometric methods can be used to carry out inference in this model are provided in the appendix. Suffice it to note here that a great advantage of staying in the framework of the state space model given by (1) and (2) is that standard methods of posterior simulation are available. In particular, Markov chain Monte Carlo (MCMC) algorithms can be used to draw the states, $\alpha = (\alpha_1', .., \alpha_T')'$. In our empirical work, we use the method of Durbin and Koopman (2002).

As discussed in the introduction, there is strong empirical evidence that volatility issues are important in many macroeconomic problems. Thus, the error covariance matrix in the measurement equation, H_t, should be allowed to vary over time. To motivate the particular specification we choose, note that the Great Moderation of the business cycle implies that it is important that the variances of macroeconomic variables should be allowed to change over time. However, many key aspects of the transmission mechanism relate to the covariances between the errors. For instance, in many models, the immediate effect of changes in monetary policy on inflation is dependent upon the correlation between the errors in the interest rate and inflation equations. Thus, it is potentially important to allow for both the error variances and covariances to change over time.

Following Primiceri (2005), we use a triangular reduction of the measurement error covariance, H_t, such that:

$$A_t H_t A_t' = \Sigma_t \Sigma_t'$$

or

$$H_t = A_t^{-1} \Sigma_t \Sigma_t' \left(A_t^{-1} \right)', \tag{3}$$

where Σ_t is a diagonal matrix with diagonal elements $\sigma_{j,t}$ for $j = 1, .., p$ and A_t is the lower triangular matrix:

$$A_t = \begin{bmatrix} 1 & 0 & \dots & . & & 0 \\ a_{21,t} & 1 & \dots & . & & . \\ . & . & \dots & . & & . \\ . & . & \dots & 1 & & 0 \\ a_{p1,t} & . & \dots & a_{p(p-1),t} & 1 \end{bmatrix}.$$

To model evolution in Σ_t and A_t we must specify additional state equations. For Σ_t a stochastic volatility framework can be used. In particular, if $\sigma_t = (\sigma_{1,t}, .., \sigma_{p,t})'$, $h_{i,t} = \ln(\sigma_{i,t})$, $h_t = (h_{1,t}, .., h_{p,t})'$ then Primiceri uses:

$$h_{t+1} = h_t + u_t, \tag{4}$$

where u_t is $N(0, W)$ and is independent over t and of ε_t and η_t. Technical details for drawing h in an MCMC algorithm are given in the appendix. Here we stress only that standard algorithms are required. In our empirical work, we use the algorithm of Kim, Shephard and Chib (1998).

To describe the manner in which A_t evolves, we first stack the unrestricted elements by rows into a $\frac{p(p-1)}{2}$ vector as $a_t = \left(a_{21,t}, a_{31,t}, a_{32,t}, .., a_{p(p-1),t}\right)'$. These are allowed to evolve according to the state equation:

$$a_{t+1} = a_t + \zeta_t, \tag{5}$$

where ζ_t is $N(0, C)$ and is independent over t and of u_t, ε_t and η_t. With regards to our MCMC algorithm, we can transform the original measurement equation so that the Durbin and Koopman (2002) algorithm can be used to draw the states.

Primiceri (2005) uses the model given by (1) through (5), which we will refer to as a TVP-VAR with stochastic volatility, in a study of the evolution of monetary policy. Cogley and Sargent (2005) use a similar specification but one which has parameters comparable to A_t being constant over time. Cogley and Sargent (2001) uses an even more restricted variant of this specification which does not have multivariate stochastic volatility (i.e. it is a TVP-VAR, but H_t is constant over time). This (very incomplete) discussion of the related literature is meant to motivate that this class of models is receiving a great deal of attention by macroeconomists. These are very flexible models which are well-suited for estimating transmission mechanisms and their evolution over time. However, they contain very many parameters and, thus, there is the risk that they will over-fit the data. A common symptom of over-fitting is if a model yields good in-sample performance, but poor out–of-sample

forecast performance. It is perhaps significant that most papers using this sort of models present only in-sample results. These considerations motivate the development of models which are flexible, but allow for a more tight parameterization to lessen the risks of over-fitting. It is to such an extension which we now turn.

Time varying parameter models imply that coefficients change every time period, although the magnitude of the change in coefficients can be restricted by the state equation. That is, typically the error covariance matrix in the state equation is estimated to be small and, thus, α_{t+1} is close to α_t. Thus, TVP models work well when the evolution of coefficients is constant but gradual. Loosely speaking, TVP models can be thought of as "many small breaks" models. In contrast to TVP models, there is a large literature which assumes that fewer changes in coefficients occur, but when a structural break occurs, the magnitude of the change in coefficients is unrestricted. Loosely speaking these can be thought of as "few large breaks" models. Examples include Chib (1998), Maheu and Gordon (2007), Pastor and Stambaugh (2001) and Pesaran, Pettenuzzo and Timmerman (2007). See the discussion in Koop and Potter (2007a) for attempts to reconcile these different approaches.

An increasingly popular class of models which are increasingly used to model structural breaks are mixture innovation models. McCulloch and Tsay (1993) is an early example of such an approach, Gerlach, Carter and Kohn (2000) develops a very efficient computational algorithm and Giordani and Kohn (2006) applies mixture innovation models to change-point problems.

The mixture innovation aspect arises by allowing some or all of the states and parameters in the previous models to be determined (up to a set of unknown parameters) by a sequence of Markov random vectors $K = (K_1, .., K_T)'$. As we shall see shortly, these vectors will control the structural breaks in the model. In our model, we allow for breaks in the VAR coefficients (α_t) and the measurement error covariance matrix (H_t). Remember that $H_t = A_t^{-1} \Sigma_t \Sigma_t' \left(A_t^{-1} \right)'$ and, thus, the measurement error covariance matrix is parameterized in terms of Σ_t and A_t. Given that some authors (e.g. Cogley and Sargent, 2005) assume a time-invariant A_t, there does seem to be interest in models with breaks in the error variances, but not covariances. Accordingly, we allow for an unknown number of breaks in α_t, Σ_t and A_t and (of empirical importance and in contrast to much of the literature on structural breaks), we allow for breaks in these three sets of parameters to occur at different times. Accordingly, we let $K_t = (K_{1t}, K_{2t}, K_{3t})'$ for $t = 1, .., T$, where $K_{1t} \in \{0, 1\}$ controls breaks in the VAR coefficients, $K_{2t} \in \{0, 1\}$

controls breaks in Σ_t and $K_{3t} \in \{0, 1\}$ controls breaks in A_t.

We extend the TVP-VAR with stochastic volatility model as follows. In (2), the state equation which controls the evolution of α_t, we set $R_t = K_{1t}$. Note that this implies that there are time periods when the VAR coefficients remain constant ($K_{1t} = 0$) and times when a break in the VAR coefficients can occur ($K_{1t} = 1$).

(4) and (5) are the state equations which control the evolution in Σ_t and A_t. We generalize these to:

$$h_{t+1} = h_t + K_{2t} u_t \tag{6}$$

and

$$a_t = a_{t-1} + K_{3t} \zeta_t. \tag{7}$$

Thus, these vectors of parameters can either remain constant ($K_{2t} = 0$ and/or $K_{3t} = 0$) or a break can occur ($K_{2t} = 1$ and/or $K_{3t} = 1$). All other assumptions given for the TVP-VAR with stochastic volatility model still hold.

Note that all of the models previously discussed are nested within this mixture innovation extension of the TVP-VAR with stochastic volatility. If $K_{1t} = K_{2t} = K_{3t} = 1$ for $t = 1, .., T$, then we obtain the TVP-VAR with stochastic volatility of Primiceri (2005). If $K_{1t} = K_{2t} = K_{3t} = 0$ for $t = 1, .., T$, then we obtain the traditional VAR with constant parameters. If $K_{1t} = 1$ and $K_{2t} = K_{3t} = 0$ then we obtain a homoskedastic TVP-VAR as in Cogley and Sargent (2001). If $K_{1t} = K_{2t} = 1$ and $K_{3t} = 0$ then we obtain the model of Cogley and Sargent (2005). Different configurations allow for change-points to occur at different times. It is also worth stressing that, unlike most other approaches to structural break modelling, the mixture innovation framework allows us to deal with the case where there is an unknown number of change-points. As discussed in Koop and Potter (2007b) this is an advantageous feature since imposing the restriction that a fixed number of breaks occur leads to models with undesirable characteristics.

To complete the model, we must specify a hierarchical prior for K. The posterior simulation algorithm for K discussed in the appendix will work provided the hierarchical prior for K_t is Markov. In our empirical work, we adopt a Bernoulli distribution:

$$p(K_{jt} = 1) = p_j \tag{8}$$

9

for $j = 1, 2, 3$. Thus, p_j is the probability that a break occurs at time t, for $j = 1, 2, 3$ (i.e. corresponding to α_t, Σ_t or A_t). This is treated as an unknown parameter and estimated from the data. In our empirical work, we allow for breaks to occur independently in α_t, Σ_t or A_t (i.e. K_{1t}, K_{2t} and K_{3t} are independent of one another, contemporaneously and at all leads and lags). But correlations between the breaks in the different blocks of parameters could easily be allowed for with trivial changes in the MCMC algorithm. This algorithm is described in detail in the Technical Appendix. Here we note only that it takes the MCMC algorithm for the TVP-VAR with stochastic volatility and adds extra steps taken from Giordani and Kohn (2006) for the mixture innovation aspect of the model.

3 Macroeconomic Issues

3.1 The Data

To investigate issues relating to monetary policy, it is common (e.g. Cogley and Sargent, 2001 and 2005, Primiceri, 2005, or Stock and Watson, 2001) to use a short term interest rate as being under the control of the Fed (the "policy block") with the inflation and unemployment rates representing the "non-policy block". Accordingly, we use data from 1953Q1 through 2006Q2 on the unemployment rate (seasonally adjusted civilian unemployment rate, all workers over age 16), interest rate (yield on three month Treasury bill rate) and inflation rate (the annual percentage change in a chain-weighted GDP price index).[2]

3.2 Features of Interest

The models described thus far are reduced form models. Identifying assumptions must be made to allow for structural interpretation. We go from our time-varying reduced form VARs to time-varying structural form VARs in a standard way (see, e.g., Primiceri, 2005) and begin by ordering our dependent variables as inflation, unemployment and interest rates in the vector y_t. In particular, in our reduced form models the errors in the measurement equation, ε_t, were $N(0, H_t)$ where H_t is parameterized as in (3). The structural

---

[2]The data were obtained from the Federal Reserve Bank of St. Louis website, http://research.stlouisfed.org/fred2/.

form errors, u_t are assumed to be $N(0, I)$ and the structural form model has

$$y_t = Z_t\alpha_t + \Upsilon_t u_t \qquad (9)$$

where Υ_t imposes the identifying restrictions. With regards to the policy block, we assume that the shock to the interest rate equation (i.e. the monetary policy shock) has no immediate effect on inflation and unemployment. This is a standard assumption used, among many others, by Bernanke and Mihov (1998), Christiano, Eichanbaum and Evans (1999) and Primiceri (2005). Identification in the non-policy block is achieved by assuming that the shock to the unemployment equation has no immediate effect on inflation.[3] These identifying assumptions imply that Υ_t is lower triangular. The relationship between the reduced form and structural form parameters thus becomes:

$$\Upsilon_t = A_t^{-1}\Sigma_t.$$

There are, of course, many macroeconomic features that can be presented with a structural VAR model such as the one discussed. However, for our policy question, the most important ones relate to monetary policy. With regards to the exogenous shocks, we simply plot their standard deviations (i.e. the diagonal elements of Σ_t) with the standard deviation of the interest rate equation being of greatest importance as reflecting the monetary policy shock. With regards to the transmission mechanism, impulse response functions are of interest. Given our interest in evolving monetary policy, we focus on the impulse response of the variables in the non-policy block (i.e. inflation and interest rates) to policy (i.e. to the monetary shock).

With nonlinear time series models such as the TVP-VARs we are working with, there are some issues which arise with impulse response analysis which do not arise with linear (time-invariant) models (see Koop, 1996, and Koop, Pesaran and Potter, 1996). These are discussed in the Technical Appendix. Suffice it to note here that, following other authors such as Primiceri (2005), we calculate impulse responses for a shock at time τ with response over any time period from τ to $\tau + n$, based on the parameters as they are at time τ.

---

[3]This assumption is more controversial since we could have assumed the inflation shock had no immediate effect on unemployment. But, as discussed in Primerci (2005), empirical results are very similar for these two assumptions.

4 Empirical Results

We divide our empirical results into two sub-section. The first of these discusses the evolution of parameters in the VAR model and, in particular, whether there is evidence for parameter change and, if so, in which parameters and of what sort. The second presents results on our macroeconomic features of interest. Throughout, we present results with two lags in the VAR and an intercept (but no additional deterministic terms).

4.1 Evidence on Parameter Evolution

Before presenting empirical results relating to macroeconomic features of interest, we present some direct evidence on whether breaks have occurred in our three blocks of parameters (i.e. the VAR coefficients, the volatilities, Σ_t, and A_t which relates to the error covariances) and, if so, of what sort. A convenient vehicle for discussing these issues is through our mixture innovation variables which control the changes in the three sets of parameters, K_1, K_2 and K_3 (or their associated transition probabilities, p_1, p_2 and p_3). As discussed in section 2 of the paper, by setting particular values for K_1, K_2 and K_3, we can obtain many different models of interest. The ones we consider are listed in Table 1. We consider various restricted versions of our model as noted in Table 1 including the models of Primiceri (2005) and of Cogley and Sargent (2005) where the latter restricts A_t to be constant over time. We also consider a homoskedastic TVP-VAR model as well as a model with multivariate stochastic volatility, but constant VAR coefficients. This latter is motivated by papers such as Sims and Zha (2006), which have found support for models with no changes in the VAR coefficients (but substantive changes in the error covariance matrix).

The prior used in the paper is described in the Technical Appendix. It is a training sample prior of the sort used by Primiceri (2005) and Cogley and Sargent (2001, 2005). Indeed, for the TVP-VAR with stochastic volatility it is the same as Primiceri's prior. The new parameters relate to the mixture innovation extension. As discussed in the appendix, we use Beta priors for p_j and, thus, $B\left(\underline{\beta}_{1j}, \underline{\beta}_{2j}\right)$ for $j = 1, 2, 3$. The properties of the Beta distribution are given, e.g., in Koop (2003, page 330) and, from these, it can be seen that if $\underline{\beta}_{11} = 1, \underline{\beta}_{2j} = 1$ for all j, then $E(p_j) = \frac{1}{2}$ (with standard deviation 0.29). This is our Benchmark prior, which says, *a priori*, that there is a 50% chance

of a break occurring in any time period. The standard deviation is very large, indicating a relatively noninformative prior.

Inspired by much of the structural break literature which works with models with a small number of breaks (e.g., among many others, Pesaran, Pettenuzzo and Timmerman, 2007), we might also be interested in working with a model which only allows for, say, one or two breaks. However, one of the advantages of the mixture innovation approach to structural break modelling is that it does not impose, *a priori*, a fixed number of breakpoints on the data. Instead it estimates the number of breakpoints in a data-based fashion. So, we cannot simply choose a mixture innovation model with, say, one or two breaks imposed. However, we can tighten the prior on the transition probabilities towards such a model. This is what we do in the model labelled "Few Breaks" in Table 1. In particular, for the prior hyperparameter values listed in Table 1 for this model we have $E(p_j) = 0.001$ (with standard deviation 0.010) for $j = 1, 2, 3$.

Table 1: Models and Priors Used in the Empirical Work			
Model	Prior or Modelling Assumptions Relating to:		
	VAR coefficients	Σ_t	A_t
Benchmark	$\underline{\beta}_{11} = 1, \underline{\beta}_{21} = 1$	$\underline{\beta}_{11} = 1, \underline{\beta}_{22} = 1$	$\underline{\beta}_{13} = 1, \underline{\beta}_{23} = 1$
Benchmark A_t constant	$\underline{\beta}_{11} = 1, \underline{\beta}_{21} = 1$	$\underline{\beta}_{11} = 1, \underline{\beta}_{21} = 1$	$K_{3t} = 0$ for all t
Benchmark A_t, Σ_t constant	$\underline{\beta}_{11} = 1, \underline{\beta}_{21} = 1$	$K_{2t} = 0$ for all t	$K_{3t} = 0$ for all t
Benchmark α_t constant	$K_{1t} = 0$ for all t	$\underline{\beta}_{11} = 1, \underline{\beta}_{21} = 1$	$\underline{\beta}_{13} = 1, \underline{\beta}_{23} = 1$
Primiceri	$K_{1t} = 1$ for all t	$K_{2t} = 1$ for all t	$K_{3t} = 1$ for all t
VAR	$K_{1t} = 0$ for all t	$K_{2t} = 0$ for all t	$K_{3t} = 0$ for all t
Few breaks	$\underline{\beta}_{11} = .01, \underline{\beta}_{2j} = 10$	$\underline{\beta}_{1j} = .01, \underline{\beta}_{2j} = 10$	$\underline{\beta}_{1j} = .01, \underline{\beta}_{2j} = 10$

Table 2 presents empirical results relating to the question of which type of model receives support from the data. An advantage of our mixture innovation approach is that evidence for or against any particular restricted version of our model (such as those listed in Table 1) can be revealed by looking at posterior of parameters such as p_1, p_2 and p_3. The usual method of Bayesian model comparison is through marginal likelihoods and (although

we do present marginal likelihoods) these can be more sensitive to prior information than posteriors (especially with models such as ours with high-dimensional parameter spaces). Thus, much of our discussion of how monetary policy evolves relates to p_1, p_2 and p_3.

In addition to the marginal likelihood, we also present the expected value of the log-likelihood function. The Technical Appendix discusses how these measures of model performance are calculated and how the expected value of the log-likelihood can be interpreted as the empirical Bayesian metric described in Carlin and Louis (2000, section 6.5.1) and are closely related to conventional information criteria. In Table 2, the column labelled "$E[\log L]$" presents this measure of model performance.

Regardless of whether we look at the posteriors for p_1, p_2 and p_3, the marginal likelihoods or the expected log likelihoods, the story that comes through is a strong one. We are finding that all three of our sets of parameters (α_t, A_t and Σ_t) do change over time and in a way that is closer to being the gradual evolution of the TVP-VAR than the abrupt breaks of conventional structural break models. Consistent with the Great Moderation of the business cycle, we are finding most evidence for evolution of the error variances. We elaborate on these points in the next paragraph.

Our mixture innovation TVP-VAR with stochastic volatility estimates all three of our transition probabilities to be above 0.8 indicating that, in any time period, there is a very high probability that parameters will change. We are thus finding support for a model that is close to Primiceri's model (although, of course, we did not impose this on the data). It is worth noting that our model has a slightly higher marginal likelihood and expected log-likelihood than Primiceri's model. Among the other restricted versions of our model, the one which imposes A_t as being constant does the best. This is the model of Cogley and Sargent (2005). The remaining models clearly are receiving little support. The standard VAR with time-invariant parameters does the worst. The TVP-VAR with constant error covariance (used in Cogley and Sargent, 2001) also does very poorly. Finally, it is worth noting that there does seem to be time variation in the VAR coefficients as the model which restricts α_t to be constant receives little support.

All of the results discussed so far use the Benchmark Prior (or suitable restricted variants of it). We have carried out a prior sensitivity analysis and results are similar, even when we make substantive changes in the prior hyperparameter values. As one example, consider the Few Breaks prior which expresses extremely strong views that the transition probabilities are near zero

14

(i.e. $E(p_j) = 0.001$ with standard deviation 0.010). Using this prior, the posteriors for the transition probabilities are $E(p_1|Data) = 0.18$, $E(p_2|Data) = 0.87$ and $E(p_3|Data) = 0.27$ (with posterior standard deviations of 0.04, 0,04 and 0.07, respectively). Thus, even though we have used prior information that the transition probabilities are near zero, data information is so strong that the posteriors are pulled strongly in directions which suggest gradual change in all parameters. Note that, in the case of α_t and A_t, we are finding that the prior does have some effect (e.g. $E(p_3|Data) = 0.82$ with the Benchmark Prior and $E(p_3|Data) = 0.27$ with the Few Breaks prior), but the point estimates still indicate gradual evolution of coefficients (i.e. even a transition probability of 0.27 indicates we would expect a break to occur about once a year). As before, the evidence of parameter change is greatest for Σ_t, but is still appreciable for α_t and A_t.

In summary, thus far we have made a strong case in favor of our mixture innovation extension of a TVP-VAR with stochastic volatility. In terms of the controversies in the macroeconomics literature, we are finding (like Primiceri, 2005) that parameter change in the error covariance matrix is of predominant importance, but that evolution in the VAR coefficients is appreciable enough that it should not be neglected. However, so far our argument has been purely statistical, using reduced form models. The question of what kind of implications this parameter evolution has for our understanding of monetary policy has to now be addressed.

Table 2: Results using Benchmark Prior for Mixture Innovation TVP-VAR with Multivariate Stochastic Volatility and Restricted Versions of this Model					
Model	Marginal Likelihood	$E(\log L)$	$E(p_1|Y)$	$E(p_2|Y)$	$E(p_3|Y)$
Benchmark	18.51	3.91	0.95 (0.05)	0.96 (0.03)	0.82 (0.15)
Benchmark A_t constant	4.34	2.66	0.96 (0.03)	0.96 (0.03)	0.00 (0.00)
Benchmark A_t, Σ_t const.	5×10^{-5}	0.15	0.94 (0.06)	0.00 (0.00)	0.00 (0.00)
Benchmark α_t constant	3×10^{-6}	-0.45	0.00 (0.00)	0.94 (0.04)	0.79 (0.16)
Primiceri	15.92	3.59	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
VAR	1×10^{-13}	-5.94	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)

4.2   Evidence on the Evolution of Monetary Policy

4.2.1   The Volatility of Exogenous Shocks

The results of the previous section clearly indicate that the parameters of the reduced form models have changed over time. There is particularly strong evidence that the error covariance matrix has changed. Accordingly, in this section we present only a brief discussion of models (such as the standard VAR) which have a constant error covariance matrix. However, as we have discussed, there are controversies in this literature over whether the monetary transmission mechanism has changed over time or whether it is merely the properties of the exogenous shocks which have changed. The results in our previous section suggest that both have changed, but evidence for the transmission mechanism changing is weaker. Given these general considerations, in this section, in addition to our mixture innovation TVP-VAR with stochastic volatility, we also consider the TVP-VAR with stochastic volatility of Primiceri (2005) as well as a restricted version of our model which has stochastic volatility, but the VAR coefficients are time-invariant (i.e. our Benchmark model with the restriction that α_t is constant over time). These are labelled, "Benchmark", "Primiceri" and "Hetero VAR", respectively, in the figures.

Before discussing impulse responses, we begin with some evidence directly

relating to the parameters of these models. Figure 1 through 3 plot the point estimates of the standard deviations of the errors in the measurement equation (i.e. the posterior mean of the square roots of the diagonal elements of H_t). These figures do indicate substantial variation in volatility.[4] The general patterns in these graphs are similar to those noted by others in the literature. For instance, our Figures 1 through 3 look very similar to Figures 1 a) through c) in Primiceri (2005). There is the same increase in inflation volatility until 1975 and, subsequently, a tendency (with many exceptions, particularly in the early 1980s) for it to decline. The volatility in unemployment equation error spikes around 1975. The volatility of the monetary policy shock from the interest rate equation shows a big increase in the early 1980s before becoming much lower afterwards (with the interesting exception of a substantial increase in 2001).

With regards to comparing the three different models listed above, they all do capture the same broad patterns of volatility for all variables. However, some interesting differences do exist. This is most noticeable in Figure 1 where the heteroskedastic VAR yields a much smoother pattern of volatility for the inflation equation. There are also noticeable divergences between our model and that of Primiceri (2005), particularly in the crucial mid-1970s through early 1980s time period where most of the change in the parameters appears to happen. For the other two equations (see Figures 2 and 3), fewer differences exist. Our mixture innovation model and the TVP-VAR with stochastic volatility are yielding quite similar patterns of volatility, but the heteroskedastic VAR diverges from this pattern a few times, particularly in the unemployment equation. However, with respect to the monetary shock (Figure 3), these three models are yielding very similar results.

---

[4]To keep the figures clear, we have not put measures of uncertainty (e.g. 10th/90th percentile bands) in these figures. These do indicate some degree of imprecision, but the change in volatility is still large relative to this imprecision.

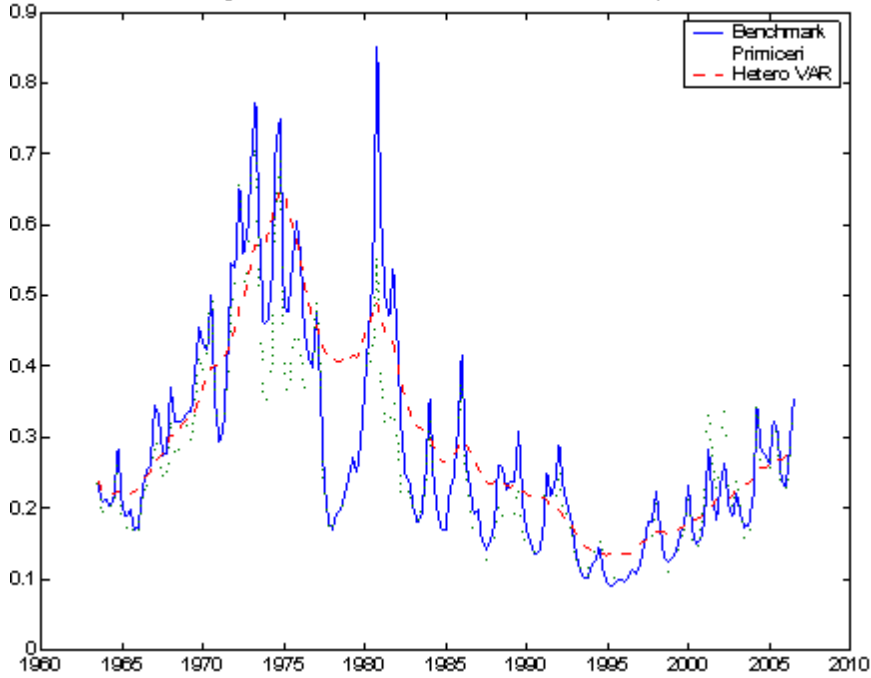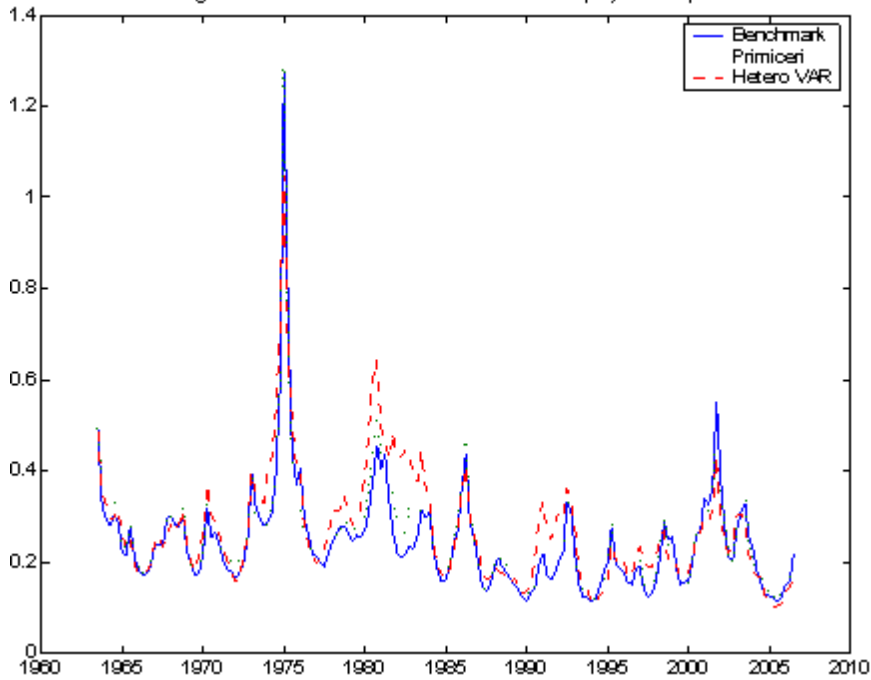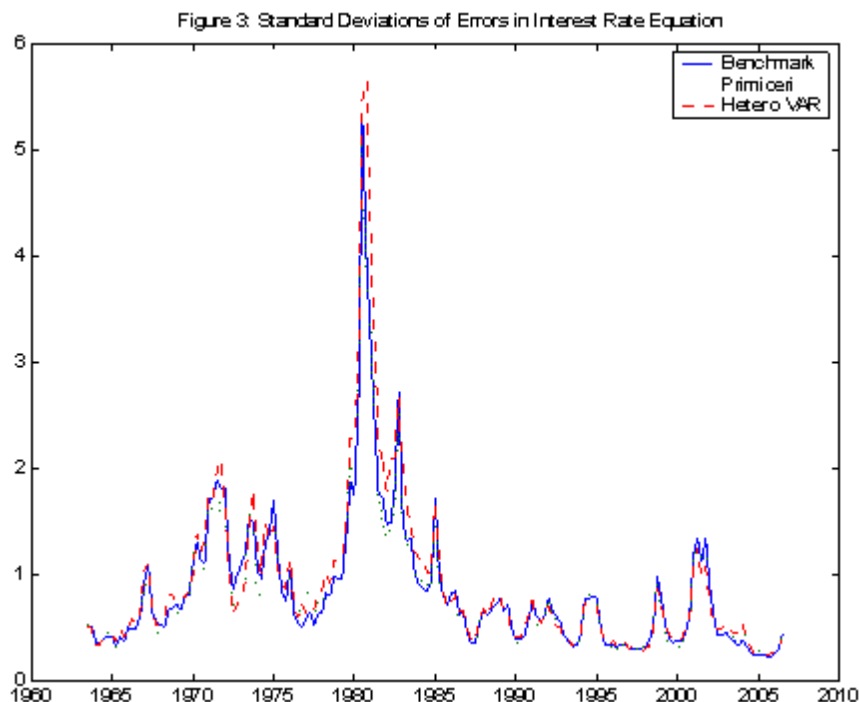Figure 1: Standard Deviations of Errors in Inflation Equation

Figure 2: Standard Deviations of Errors in Unemployment Equation

18

Figure 3: Standard Deviations of Errors in Interest Rate Equation

4.2.2 The Parameters

Given the time-varying nature of the parameters, there are too many to easily present. However, in the Empirical Appendix, we present coefficient estimates for Υ_t (which imposes the structural identification assumptions) and the VAR coefficients (α_t) for four time periods: 1975Q1, 1981Q3, 1996Q1 and 2006Q3. The first three of these are those used in Primiceri (2005) as representing a wide range of different time periods and economic conditions. The fourth date is the last in our data set. The interested reader is left to peruse Table A.1 and A2 in the Empirical Appendix in detail. A general pattern is that the coefficients are often imprecisely estimated. This is not surprising in a model such as this one with many parameters. However, keeping in mind this imprecision, it does seem that there is much greater change in Υ_t than the VAR coefficients. Note, in particular, the large drop in the standard deviation of the error in the unemployment equation (consistent with the Great Moderation of the business cycle). Also some of the parameters which determine the correlations between the errors have changed greatly. Υ_{1996Q1} and Υ_{2006Q3} look quite similar to one another. And, for many elements, Υ_{1981Q3}

19

also looks the same. However, there are exceptions where some elements of Υ_{1981Q3} are quite different from other years. For Υ_{1975Q1}, some elements are very different from later years (but others are quite similar). All in all, the general pattern is of different elements of Υ_t changing at different times (but some elements not changing much at all) with most (but not all) of the change occurring prior to 1996.

With the VAR coefficients (Table A.1) there is less evidence of parameter change (relative to the level of imprecision in the posteriors). However, there does seem to be some moderate movement in some coefficients and our evidence from the previous section indicate change is occurring. Interestingly, more of the parameter change seems to occur near the end of the sample (not in the 1970s and 1980s as with the error covariance matrix). A key question is whether this change has important macroeconomic implications and this is a question best investigated through impulse response functions.

Finally, we note briefly that Tables A1. and A.2 present results for a standard homoskedastic VAR. These are often quite different from TVP-VAR results. In particular, it is worth noting that the homoskedasticity assumption causes the VAR coefficients to be far off in some cases. This suggests that, even if interest only centers on VAR coefficients, it is not safe to ignore heteroskedasticity.

4.2.3 Impulse Response Functions: Comparing Different Time Periods

Given our interest in the evolution of monetary policy, we focus on impulse responses to a monetary shock. To be precise, we calculate the effect of a shock of size one to the structural errors (i.e. using equation 9, we set the third element of u_t to one and trace out its effect on all the variables). Figures 4, 5 and 6 present point estimates (posterior medians) of the impulse response to inflation, unemployment and the interest rate for our four representative time periods: 1975Q1, 1981Q3, 1996Q1 and 2006Q3.

The response of the interest rate to the monetary shock (see Figure 6) is very nearly the same in every time period, but the responses of inflation and the unemployment rate exhibit more interesting patterns. The point estimate of the response of inflation to the monetary shock (Figure 4) in 1975Q1 is very different from all the other years. At short and medium horizons, the pattern is that monetary shocks are having less of an effect as time goes by. In 1975Q1 there is a large (positive) hump-shaped response of inflation to a monetary

shock. But in later years this sort of response has vanished. The point estimate of the response of the unemployment rate to the monetary shock also shows a similar pattern, but the vanishing hump occurs later. That is, there is a large (positive) hump-shaped response of unemployment to a monetary shock (at short and medium horizons) for both 1975Q1 and 1981Q3, but this, to all intents and purposes, vanishes for 1996Q1 and 2006Q3.
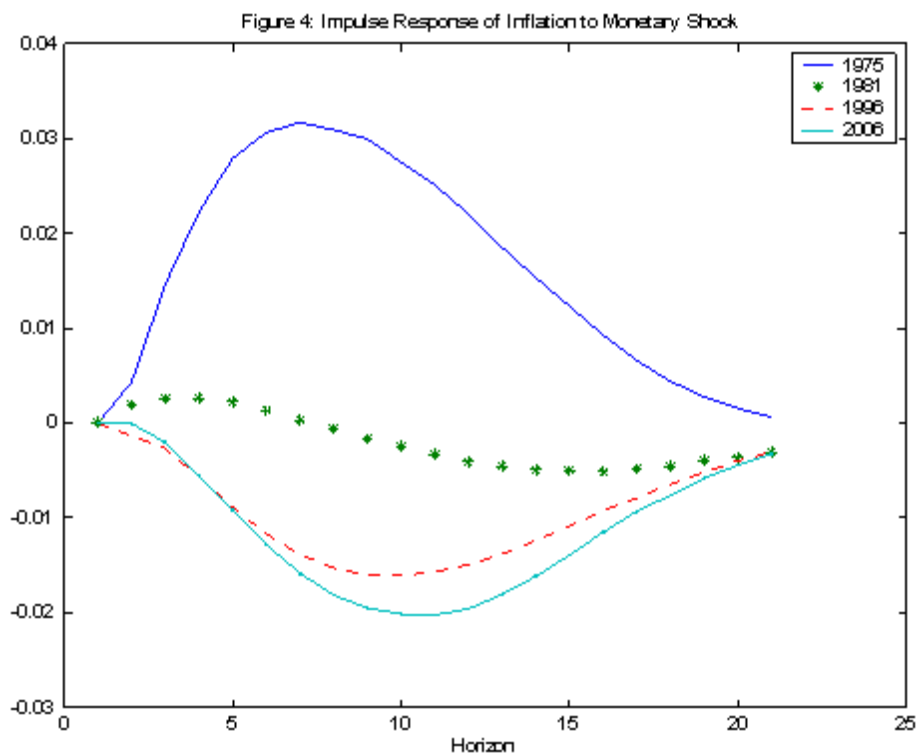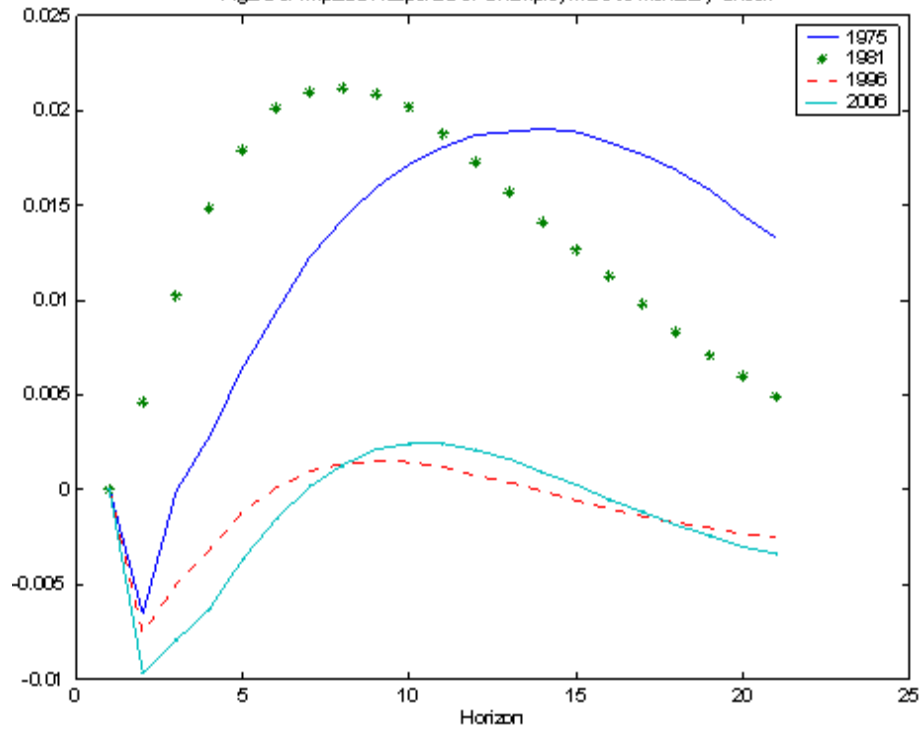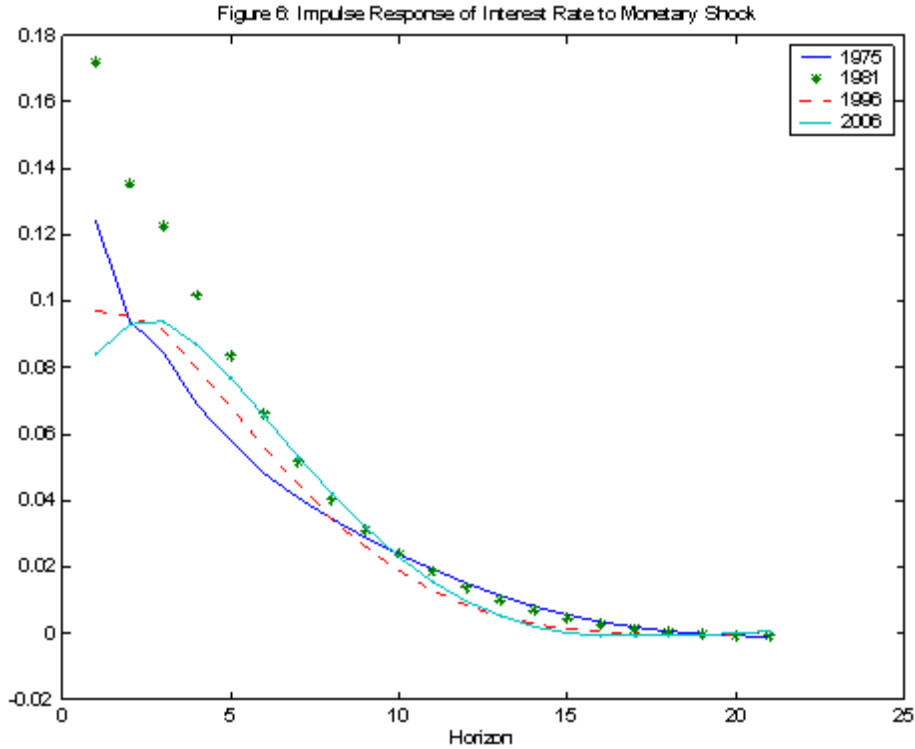

Figure 4: Impulse Response of Inflation to Monetary Shock

Figure 5: Impulse Response of Unemployment to Monetary Shock

Figure 6: Impulse Response of Interest Rate to Monetary Shock

The preceding statements about the evolution of the effects of monetary shocks were based on point estimates of impulse responses. To keep the figures readable, we did not put measures of uncertainty associated with the point estimates. These tend to be quite large. In Figure 4, the point estimate of the impulse response in 1975Q1 is quite different from 1981Q3. Figure 7 relates to the difference in these impulse responses between these two years. It plots the point estimate of this difference along with the 10th and 90th percentiles of the posterior. Other years and other differences between impulse responses function exhibit a similar pattern. The basic pattern is that a horizontal line at zero always lies between the 10th and 90th percentiles (i.e. lies within our 80% credible interval). These credible intervals are calculated pointwise. Thus, the fact that each credible interval contains zero individually does not necessarily imply that jointly there are no interesting differences in impulse responses over time. However, our previous discussion about the systematic patterns in the evolution of impulse responses must be qualified due to the inaccuracy of point estimates.
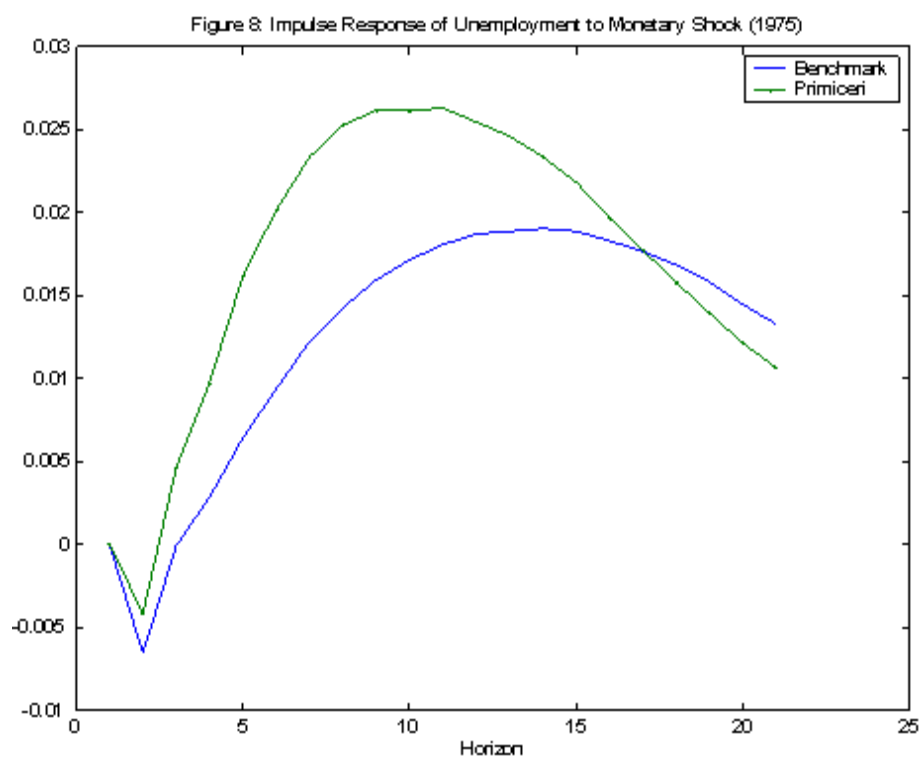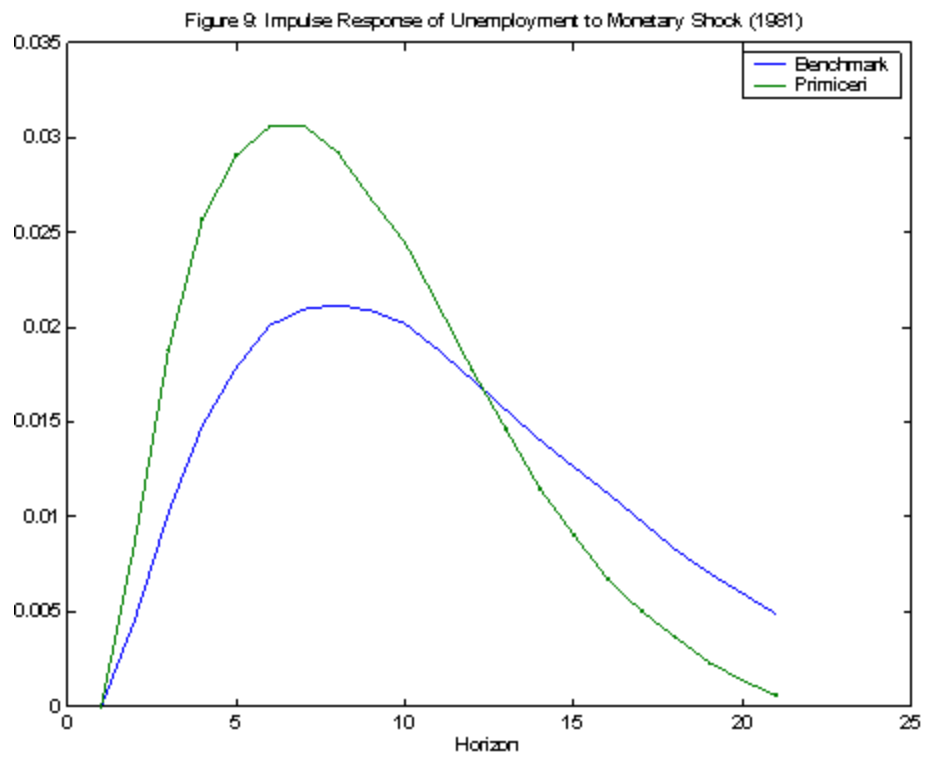
23

Figure 7: Diff. Between Impulse Responses, 1975-1981 (Resp. of Infl. to Monetary Shock)
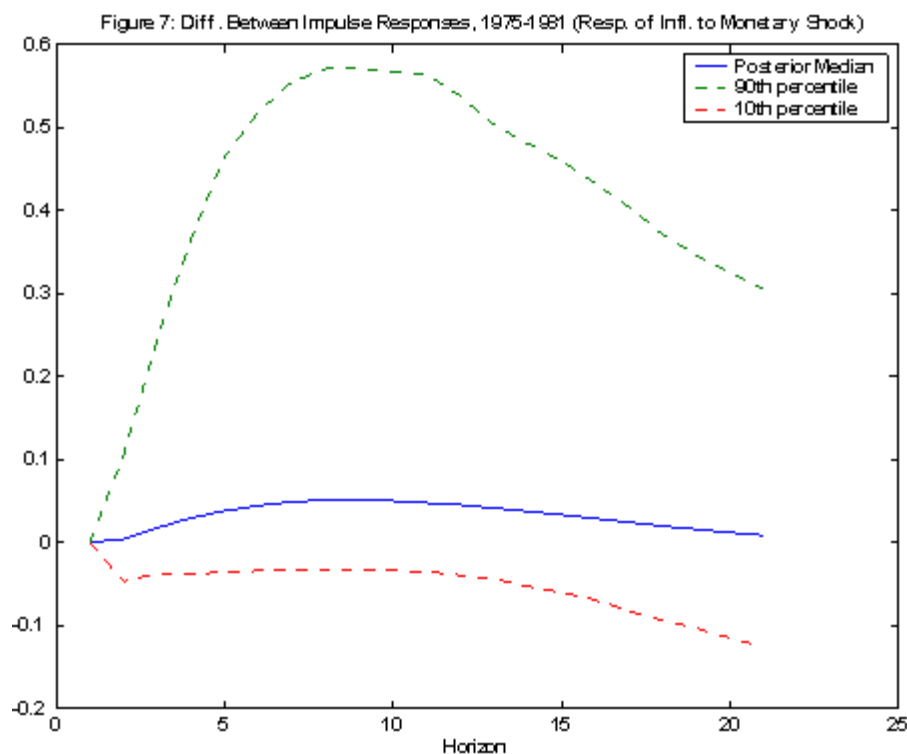
4.2.4 Impulse Responses: Comparing Different Models

For the reasons discussed on the section of the "Volatility of Exogenous Shocks", it is of most interest to compare the models we have been calling Benchmark, Primiceri and heteroskedastic VAR. For the sake of brevity, we do not plot all of the impulse responses, but rather choose the response of unemployment to the monetary shock. The other impulse responses show qualitatively similar patterns. The impulses response functions for the Heteroskedastic VAR are not plotted since even the posterior median exhibits explosive behavior. Remember that we are not ruling out explosive behavior for our VARs. For our Benchmark model and the TVP-VAR, there is little posterior probability associated with such behavior. But with the Heteroskedastic VAR and a standard VAR, there is much more evidence of non-stationarity.

Figures 8 through 11 compare results for our Benchmark model and the model of Primiceri (2005) for our four different representative time periods. It can be seen that, in all of our time periods, the two models are giving

basically the same story. However, the fact that we are unable to even plot
the heteroskedastic VAR results on a figure with the same scaling implies
that the story of our model is quite different from the heteroskedastic VAR.

Figure 8: Impulse Response of Unemployment to Monetary Shock (1975)

Figure 9: Impulse Response of Unemployment to Monetary Shock (1981)

Figure 10: Impulse Response of Unemployment to Monetary Shock (1996)

Figure 11: Impulse Response of Unemployment to Monetary Shock (2006)
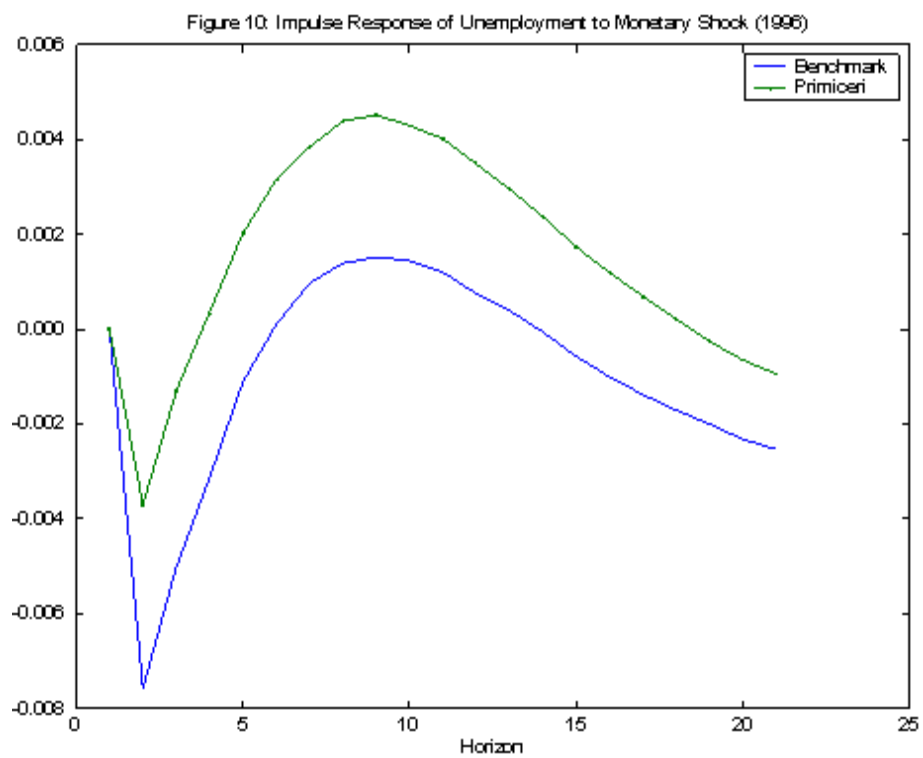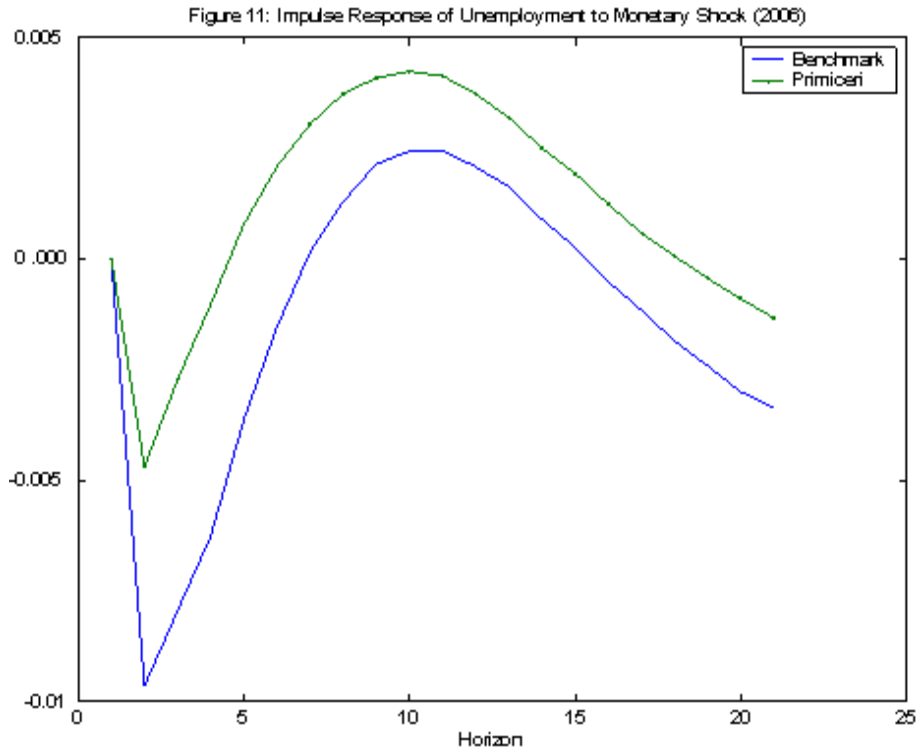
5 Conclusion

There has been much interest in the recent macroeconomic literature on the transmission of monetary policy shocks and the volatility of such shocks. In particular, questions arise over whether they have changed over time and, if so, of what form the change takes (e.g. has the change been gradual or abrupt). In this paper, we develop a model which allows us to directly address such issues. Instead of estimating a model which assumes a particular form of parameter change (i.e. TVP models assume gradual evolution of parameters whereas conventional structural break models assume a small number of abrupt breaks), our model allows us to estimate the type of change which is occurring.

We use our model in an empirical context involving three standard variables: inflation, unemployment and interest rates and make standard identifying assumptions. Our results are clear. Relative to a traditional (homoskedastic) VAR, there is overwhelming evidence of parameter change. The

strongest change relates to the error variances, but there also seems to be appreciable change in VAR coefficients and error covariances. Furthermore, this change is gradual as opposed to being abrupt.

Relative to the existing literature, our model is yielding results which are quite close to those of Primiceri (2005). Our model is also yielding results which are close (but not quite so close) to those obtained from the model of Cogley and Sargent (2005). The restrictions on the error covariance matrix in the latter paper do not receive statistical support, but freeing up the restrictions has only small economic implications. Using a different modeling strategy, Sims and Zha (2006) find evidence in support of a model where error variances and covariances change, but VAR coefficients do not. In the framework of our model, restricting VAR coefficients to be constant has only minor implications for the volatility of exogenous shocks, but has a substantive impact on impulse response functions.

References

Bernanke, B. and Mihov, I., 1998, Measuring monetary policy, *Quarterly Journal of Economics*, 113, 869-902.

Boivin, J. and Giannoni, M., 2006, Has monetary policy become more effective? *Review of Economics and Statistics*, 88, 445-462.

Carlin, B. and Louis, T., 2000, Bayes and Empirical Bayes Methods for Data Analysis, second edition (Chapman and Hall, Boca Raton).

Chib, S., 1998, Estimation and comparison of multiple change-point models, *Journal of Econometrics,* 86, 221-241.

Christiano, L., Eichenbaum, M. and Evans, C., 1999, Monetary shocks: What have we learned and to what end?, pages 65-148 in J. Taylor and M. Woodford (eds.), *Handbook of Macroeconomics*, vol. 1A (Elsevier, New York).

Cogley, T. and Sargent, T., 2001, Evolving post-World War II inflation dynamics, *NBER Macroeconomic Annual*, 16, 331-373.

Cogley, T. and Sargent, T., 2005, Drifts and volatilities: Monetary policies and outcomes in the post WWII U.S, *Review of Economic Dynamics*, 8, 262-302.

Cogley, T., Morozov, S. and Sargent, T., 2005, Bayesian fan charts for U.K. inflation: Forecasting and sources of uncertainty in an evolving monetary system, *Journal of Economic Dynamics and Control*, 29, 1893-1925.

Durbin, J. and Koopman, S., 2002, A simple and efficient simulation smoother for state space time series analysis, *Biometrika*, 89, 603-616.

Gelfand, A. and Dey, D., 1994, Bayesian model choice: Asymptotics and exact calculations, *Journal of the Royal Statistical Society Series B*, 56, 501-514.

George, E., Sun, D. and Ni, S., 2006, Bayesian stochastic search for VAR model restrictions, manuscript available at http://www-stat.wharton.upenn.edu/~edgeorge/.

Gerlach, R., Carter, C and Kohn, E., 2000, Efficient Bayesian inference in dynamic mixture models, *Journal of the American Statistical Association*, 95, 819-828.

Giordani, P. and Kohn, R., 2006, Efficient Bayesian inference for multiple change-point and mixture innovation models, *Journal of Business and Economic Statistics*, forthcoming.

Kim, S., Shephard, N. and Chib, S., 1998, Stochastic volatility: likelihood inference and comparison with ARCH models, *Review of Economic Studies*, 65, 361-93.

Koop, G., 1996, Parameter uncertainty and impulse response analysis, *Journal of Econometrics*, 72, 135-149.

Koop, G., 2003, *Bayesian Econometrics* (Wiley, Chichester).

Koop, G., Pesaran, M.H. and Potter, S., 1996, Impulse response analysis in nonlinear multivariate models, *Journal of Econometrics*, 74, 77-118.

Koop, G. and Potter, S., 2007a, Estimation and forecasting in models with multiple breaks, *Review of Economic Studies*, forthcoming.

Koop, G. and Potter, S., 2007b, Prior elicitation in multiple change-point models, working paper available at http://personal.strath.ac.uk/gary.koop/.

Lubik, T. and Schorfheide, F., 2004, Testing for indeterminacy: An application to U.S. monetary policy, *American Economic Review*, 94, 190-217.

McCulloch, R. and Tsay, E., 1993, Bayesian inference and prediction for mean and variance shifts in autoregressive time series, *Journal of the American Statistical Association*, 88, 968-978.

Maheu, J. and Gordon, S., 2007, Learning, forecasting and structural breaks, manuscript available at http://www.chass.utoronto.ca/~jmaheu/mg.pdf.

Pastor, L. and Stambaugh, R., 2001, The equity premium and structural breaks, *Journal of Finance*, 56, 1207-1239.

Pesaran, M.H., Pettenuzzo, D. and Timmerman, A., 2007, Forecasting time series subject to multiple structural breaks, *Review of Economic Studies*, forthcoming.

Primiceri. G., 2005, Time varying structural vector autoregressions and monetary policy, *Review of Economic Studies*, 72, 821-852.

Sims, C. and Zha, T., 2006, Were there regime switches in macroeconomic policy? *American Economic Review*, 96, 54-81.

Stock, J. and Watson, M., 2001, Vector autoregression, *Journal of Economic Perspectives*, 15, 101-116.

Technical Appendix: Posterior Computation, Prior and Impulse Response Analysis

The models used in this paper all begin with the TVP-VAR given in (1) and (2) in the body of the text. A key step in any of our MCMC algorithms will be to draw the states, $\alpha = (\alpha'_1, .., \alpha'_T)'$. For known values of H_t, Q_t and R_t, this can be done using any of the standard algorithms for state space models. We use the algorithm of Durbin and Koopman (2002) which (for the reasons given in that paper) is more efficient than other popular alternatives. State space algorithms such as this require a treatment of the initial condition α_1. We do this by writing (1) as:

$$y_t = Z_t \alpha_0 + Z_t \alpha_t + \varepsilon_t$$

and then initializing the algorithm for drawing states by setting $\alpha_1 = 0$. Note that α_0 can be interpreted as benchmark VAR coefficients and the state equation as capturing deviations from this benchmark. The case where $R_t = 0$ for $t = 1, .., T$ then produces the standard VAR with time-invariant parameters.

Our MCMC algorithm involves drawing from the posterior of α_0 conditional on the states and other model parameters. This is straightforward since we can re-arrange the previous equation as:

$$y_t - Z_t \alpha_t = Z_t \alpha_0 + \varepsilon_t$$

and standard results for the multivariate Normal regression model (see, e.g., Koop, 2005, pages 140-141) can be used with $y_t - Z_t \alpha_t$ as the dependent variable. In our models with stochastic volatility, we also use the Durbin and Koopman (2002) algorithm for the elements relating to the measurement error covariance matrix. In these cases, we treat initial conditions in the same manner.

Our MCMC algorithms involve cycling through the full posterior conditional distributions. For simplicity, we do not list all the conditioning arguments. But we stress that all of the posteriors noted below (which are labelled as being conditional on "Data") are the full conditionals required to set up a valid MCMC algorithm.

Model 1: A Time Varying VAR with Constant Error Covariance

A simple benchmark model is the time-varying VAR with constant error covariance matrix. This is obtained by using (1) and (2) with $H_t = H, Q_t = Q$

and $R_t = I_m$. We follow the common practice of using Wishart priors for the error precision matrices in the measurement and state equations:

$$H^{-1} \sim W\left(\underline{\nu}_H, \underline{H}^{-1}\right) \tag{A.1}$$

and

$$Q^{-1} \sim W\left(\underline{\nu}_Q, \underline{Q}^{-1}\right) \tag{A.2}$$

The posterior for H^{-1} (conditional on the states) is Wishart:

$$H^{-1}|Data \sim W\left(\overline{\nu}_H, \overline{H}^{-1}\right) \tag{A.3}$$

where

$$\overline{\nu}_H = T + \underline{\nu}_H$$

and

$$\overline{H}^{-1} = \left[\underline{H} + \sum_{t=1}^{T} (y_t - Z_t\alpha_t)(y_t - Z_t\alpha_t)'\right]^{-1}.$$

The posterior for Q^{-1} (conditional on the states) is Wishart:

$$Q^{-1}|Data \sim W\left(\overline{\nu}_Q, \overline{Q}^{-1}\right) \tag{A.4}$$

where

$$\overline{\nu}_Q = T + \underline{\nu}_Q$$

and

$$\overline{Q}^{-1} = \left[\underline{Q} + \sum_{t=1}^{T} (\alpha_{t+1} - \alpha_t)(\alpha_{t+1} - \alpha_t)'\right]^{-1}.$$

A posterior simulator for this model involves drawing the states using the algorithm of Durbin and Koopman (2002) and drawing the other model parameters from (A.3) and (A.4).

Model 2: A Time Varying VAR with Stochastic Volatility

Following the argument in the body of the paper, it is probably unreasonable to assume the error covariances are constant over time. We use a

33

triangular reduction of the measurement error covariance, H_t, given in (3) with evolution of the parameters given by (4) and (5).

To carry out posterior simulation of $h = (h_1', .., h_T')'$ (conditional on α and the parameters of the model) we can, following Primiceri (2005), adapt an algorithm of Kim, Shephard and Chib (1998) as follows. Using (3) we can transform (1) as:

$$y_t^* = A_t \left(y_t - Z_t \alpha_t \right),$$

where $var\left(y_t^* \right) = \Sigma_t \Sigma_t'$ which is a diagonal matrix. Let $y_{j,t}^*$ for $j = 1, .., p$ denote the jth element of y_t^*, $y_{j,t}^{**} = \ln \left[\left(y_{j,t}^* \right)^2 + c \right]$ and $y_t^{**} = \left(y_{1,t}^{**}, .., y_{p,t}^{**} \right)'$. Note that c is referred to as an offset constant which has no effect on the following theoretical derivations. Following standard practice we set $c = 0.001$.

We can now write our specification for Σ_t as a state space model with measurement equation given by

$$y_t^{**} = 2h_t + e_t \tag{A.5}$$

and state equation (4).[5] The only problem with using standard state space algorithms is that e_t is not Normally distributed. Note, however, that since $y_{j,t}^*$ and $y_{i,t}^*$ are independent of one another (for $i \neq j$), this independence property will carry over to $e_t = (e_{1t}, .., e_{pt})'$. Thus, we can draw on the univariate results of Kim, Shephard and Chib (1998) as relating to e_{jt}. Although e_{jt} is not Normal, Kim, Shephard and Chib (1998) show how its distribution can be approximated to an extremely high degree of accuracy by a mixture of seven Normals with means and variances given in their Table 4. If $S_{jt} \in \{1, 2, 3, .., 7\}$ denotes which of the seven Normals e_{jt} is drawn from, we can construct $S_j = (S_{j1}, .., S_{jT})'$ and $S = \left(S_1', .., S_p' \right)'$ as component indicators for all elements of e_t. Conditional on S (and α and other parameters), (A.5) and (4) is a Normal linear state space model and, hence, we can use the algorithm of Durbin and Koopman (2002) to draw h_t.

The strategy above requires that we draw from the posterior of S conditional on the model parameters and states. Kim, Shephard and Chib (1998) derive the appropriate posterior conditional. Let q_i, m_i and v_i^2 for $i = 1, .., 7$

---

[5] We treat the initial conditions as in Primiceri (2005) by drawing from the training sample prior.

34

be the component probability, mean and variance of each of the components in the Normal mixture (obtained from their Table 4). Then

$$\Pr\left(S_{it} = j | Data, h_{i,t}\right) \propto q_j f_N\left(y_{i,t}^{**} | 2h_{i,t} + m_j - 1.2704, v_j^2\right) \qquad (A.6)$$

for $j = 1, .., 7$, $i = 1, .., p$ and $t = 1, .., T$.

To complete the description of the MCMC algorithm relating to Σ_t, we need to work out the conditional posterior for W (where W is defined after equation 4). We use a Wishart prior for W^{-1}:

$$W^{-1} \sim W\left(\underline{\nu}_W, \underline{W}^{-1}\right). \qquad (A.7)$$

The posterior for W^{-1} (conditional on the states) is then Wishart:

$$W^{-1} | Data \sim W\left(\overline{\nu}_W, \overline{W}^{-1}\right) \qquad (A.8)$$

where

$$\overline{\nu}_W = T + \underline{\nu}_W$$

and d

$$\overline{W}^{-1} = \left[\underline{W} + \sum_{t=1}^{T}\left(h_{t+1} - h_t\right)\left(h_{t+1} - h_t\right)'\right]^{-1}.$$

Thus, to handle stochastic volatility in Σ_t, we add to the MCMC algorithm for Model 1 steps which draw h using the state space model (A.5) and (4), S using (A.6) and W using (A.8).

Next we describe an algorithm for drawing from A_t, the unrestricted elements of which we stack by rows into a $\frac{p(p-1)}{2}$ vector as $a_t = \left(a_{21,t}, a_{31,t}, a_{32,t}, .., a_{p(p-1),t}\right)'$. These are allowed to evolve according to the state equation (5). We can transform the original measurement equation so that the Durbin and Koopman (2002) algorithm can be used to draw the states. This can be done as follows. Define

$$\widehat{y}_t = y_t - Z_t\alpha_t$$

and:

$$A_t\widehat{y}_t = \xi_t$$

35

where ξ_t is independent $N(0, \Sigma_t \Sigma_t)$ (and independent of ζ_t). We can use the structure of A_t to isolate $\widehat{y}_t$ on the left hand side and write:

$$\widehat{y}_t = C_t a_t + \xi_t. \tag{A.9}$$

Primiceri (2005), page 845 gives a general definition of C_t. For our empirical work we have $p = 3$ and, for this case,

$$C_t = \begin{bmatrix} 0 & 0 & 0 \\ -\widehat{y}_{1,t} & 0 & 0 \\ 0 & -\widehat{y}_{1,t} & -\widehat{y}_{2,t} \end{bmatrix},$$

where $\widehat{y}_{i,t}$ is the ith element of $\widehat{y}_t$. (A.9) and (5) is now in form of the state space model given in (1) and (2) and the algorithm of Durbin and Koopman (2002) can be used to draw a_t for $t = 1, .., T$.

Recall that the error ζ_t in the state equation (5) has distribution $N(0, C)$. To complete the description of the MCMC algorithm relating to A_t, we need to work out the conditional posterior for C. We use a Wishart prior for C^{-1}:

$$C^{-1} \sim W\left(\underline{\nu}_C, \underline{C}^{-1}\right). \tag{A.10}$$

The posterior for C^{-1} (conditional on the states) is then Wishart:

$$C^{-1} | Data \sim W\left(\overline{\nu}_C, \overline{C}^{-1}\right) \tag{A.11}$$

where

$$\overline{\nu}_C = T + \underline{\nu}_C$$

and

$$\overline{C}^{-1} = \left[\underline{C} + \sum_{t=1}^{T} (a_{t+1} - a_t)(a_{t+1} - a_t)'\right]^{-1}.$$

To summarize, to handle the variation in A_t, we add to the MCMC algorithm, steps which draw a_t (for $t = 1, .., T$) using the state space model (A.9) and (5), and C using (A.11). To obtain draws of get the structural VAR (see equation 9), we can use the transformation $\Upsilon_t = A_t^{-1} \Sigma_t$.

Model 3: A Mixture Innovation Time-varying VAR with Stochastic Volatility

Our mixture innovation extension of the TVP-VAR with stochastic volatility is given in (6) through (8). (8) defines a hierarchical prior which depends on the parameters p_j for $j = 1, 2, 3$. We use a (conditionally) conjugate Beta prior for p_j for $j = 1, 2, 3$: $B\left(\underline{\beta}_{1j}, \underline{\beta}_{2j}\right)$. With this choice, the conditional posterior for the breakpoint probabilities used in our MCMC algorithm is:

$$B\left(\overline{\beta}_{1j}, \overline{\beta}_{2j}\right), \tag{A.12}$$

where

$$\overline{\beta}_{1j} = \underline{\beta}_{1j} + \sum_{t=1}^{T} K_{jt}$$

and

$$\overline{\beta}_{2j} = \underline{\beta}_{2j} + T - \sum_{t=1}^{T} K_{jt}.$$

The MCMC algorithm for the time-varying parameter model (set out previously in this appendix) still, with one minor alteration, works (except now that the formulae set out above are additionally conditional on K). The alteration is that the degrees of freedom parameters, $\overline{\nu}_Q$, $\overline{\nu}_W$ and $\overline{\nu}_C$ all have T in their formulae which should be changed to $\sum_{t=1}^{T} K_{1t}$, $\sum_{t=1}^{T} K_{2t}$ and $\sum_{t=1}^{T} K_{3t}$, respectively.

To complete our MCMC algorithm, we must specify a way of drawing K. The posterior for K conditional on the states takes a simple form. This motivated some early authors (e.g. McCulloch and Tsay, 1993) to draw from K conditional on the states. However, Gerlach, Carter and Kohn (2000) point out some limitations of such a strategy. Most importantly it can be extremely inefficient since the states and K can be very highly correlated with one another. They develop an algorithm which integrates out the states analytically and draws from $p\left(K_t | Data, K_{(-t)}\right)$ where $K_{(-t)}$ denotes all the elements of K except for K_t. For state space models, Gerlach, Carter and Kohn (2000) use notation $x^{s,t}$ for all observations from s to t on any variable, x, and show that:

$$p\left(K_t | Data, K_{(-t)}\right) \propto p\left(y^{t+1,T} | y^{1,t}, K\right) p\left(y_t | y^{1,t-1}, K^{1,t}\right) p\left(K_t | K_{(-t)}\right). \tag{A.13}$$

The term $p\left(K_t|K_{(-t)}\right)$ is simply the hierarchical prior and, thus, easy to draw from. Gerlach, Carter and Kohn (2000, pages 820-822) set out an efficient algorithm for drawing from the other terms $p\left(y^{t+1,T}|y^{1,t}, K\right)$ and $p\left(y_t|y^{1,t-1}, K^{1,t}\right)$.

As discussed in Giordani and Kohn (2006), we can draw K_{1t}, K_{2t} and K_{3t} separately from one another in the context of the three state space algorithms which make up the blocks of the MCMC algorithm for time varying parameter model with stochastic volatility. Formally, this amounts to drawing from $p\left(K_{1t}|Data, K_{(-t)}, K_{2t}, K_{3t}\right)$, $p\left(K_{2t}|Data, K_{(-t)}, K_{1t}, K_{3t}\right)$ and $p\left(K_{1t}|Data, K_{(-t)}, K_{1t}, K_{2t}\right)$. That is, drawing α_t in the time varying parameter model involves use of the algorithm of Durbin and Koopman (2002) conditional on all the model parameters including H_t (see our discussion of Model 1). K_{2t} and K_{3t} are used in the definition of H_t in Model 3. Thus, the algorithm of Gerlach, Carter and Kohn (2000) can be combined with Durbin and Koopman (2002) to draw from K_{1t} and the VAR coefficients (conditional on all other model parameters including K_{2t} and K_{3t}). Similarly, the algorithm of Gerlach, Carter and Kohn (2000) can be combined with Durbin and Koopman (2002) to draw from K_{3t} and A_t (conditional on all other model parameters including K_{1t} and K_{2t}). Finally, the algorithm of Gerlach, Carter and Kohn (2000) can be combined with our extension of Kim, Shephard and Chib (1998) to draw from K_{2t} and Σ_t (conditional on all other model parameters including K_{1t} and K_{3t}).

For the TVP-VAR the prior we use is the same as that used in Primiceri (2005). That is, we use a training sample prior with the first ten years of data to choose many of the key prior hyperparameters. To be precise, we use the training sample and a time-invariant VAR to produce OLS estimates of the VAR coefficients, $\widehat{\alpha}_0$, and the error covariance matrix, $\widehat{\Omega}$ and decompose the latter as in (3) to produce $\widehat{a}_0$ and $\widehat{h}_0$ (where these are both vectors stacking the free elements as we did with A_t and Σ_t). We also obtain OLS estimates of the variance-covariance matrices of $\widehat{\alpha}_0$ and $\widehat{a}_0$ which we label $\widehat{V}_\alpha$ and $\widehat{V}_a$. Using these, we construct the priors for the initial conditions in each of our state equations as:

$$\alpha_0 \sim N\left(\widehat{\alpha}_0, 4\widehat{V}_\alpha\right),$$

$$a_0 \sim N\left(\widehat{a}_0, 4\widehat{V}_a\right)$$

and

$$\log(h_0) \sim N\left(\log\left(\widehat{h}_0\right), I_3\right).$$

Next we describe the priors for the error variances in the state equations. Note that we are choosing small degrees of freedom parameters (relative to sample size) and, thus, these prior contain a relatively small amount of information (relative to the data). For (A.2) we set $\underline{\nu}_Q = 40$ and $\underline{Q} = 0.0001\widehat{V}_\alpha$ For (A.7) we set $\underline{\nu}_W = 4$ and $\underline{W} = 0.0001I_3$. For (A.10), we set $\underline{\nu}_C = 3$ and $\underline{C} = 0.01\widehat{V}_a$.

For the TVP-VAR this completes the specification of the prior. For restricted versions of this model (e.g. the homoskedastic TVP-VAR or the standard time-invariant VAR) we use the same prior for the parameters which are left unrestricted.

The preceding prior choices were the same as Primiceri (2005) and were calibrated with the TVP-VAR with stochastic volatility in mind. With our mixture innovation extension of the TVP-VAR, we have to additionally elicit the prior hyperparameters $\underline{\beta}_{1j}$ and $\underline{\beta}_{2j}$. These are discussed in the empirical section. With regards to the remaining parameters, we make one alteration on Primiceri's prior. The latter was a prior calculated for a TVP-VAR with stochastic volatility which assumed a structural break occurred in every time period (a "many small breaks" model). We want our prior for the mixture innovation extension to allow for this, but also to allow for fewer breaks, potentially of a larger magnitude. Accordingly, we allow the mean of the error covariance matrices for the state equation to depend on our prior about the number of breaks which occur. Note that, the Beta prior in (A.12) implies that

$$E(p_j) = \frac{\underline{\beta}_{1j}}{\underline{\beta}_{1j} + \underline{\beta}_{2j}}.$$

If we let $T_{0j} = E(p_j)T$, we modify our previous prior hyperparameters as $\underline{Q} = 0.0001\widehat{V}_\alpha \frac{T}{T_{01}}$, $\underline{W} = 0.0001I_3\frac{T}{T_{02}}$ and $\underline{C} = 0.01\widehat{V}_a\frac{T}{T_{03}}$. Thus, if we set $\overline{E}(p_j) = 1$ we get Primiceri's prior, but if we use a prior for p_j which implies fewer breaks, then our prior for the state equation error variances allows for large shifts in the parameters to occur.

The Gerlach, Carter and Kohn (2000) algorithm allows us to calculate the marginal likelihood and the expected value of the likelihood in a straightforward manner. Let Y stack all the data on the dependent variables and

λ denote all the parameters in the model except for K_1, K_2 and K_3. Equation (3) and Lemmas 3 and 4 of Gerlach, Carter and Kohn (2000) describe how we can calculate $p(Y|K_1)$ for the model studied in that paper. Our algorithm uses the Gerlach, Carter and Kohn (2000) three times (i.e. for K_1, K_2 and K_3). But we can use the Gerlach, Carter and Kohn (2005) result as holding for $p(Y|K_1, K_2, K_3, \lambda)$. By averaging over MCMC draws of all of these parameters (i.e. K_1, K_2, K_3, λ), we can obtain the expected value of the log-likelihood function. To calculate the marginal likelihood, we use these draws of the likelihood function in the approach to marginal likelihood calculation of Gelfand and Dey (1994). Note that this approach involves integrating out the states before calculating $p(Y|K)$ (and, hence, is much more computationally efficient than using the Kalman filter to evaluate the likelihood function). Finally, note that some of the models set elements of K to particular values and, for these, we simply condition on these values. For instance, for the Benchmark model with α_t being constant, we calculate $p(Y|K_{11} = K_{12} = .. = K_{1T} = 0)$.

The use of the expected log-likelihood can be motivated as in Section 6.5.1 of Carlin and Louis (2000). Note that Carlin and Louis's penalized likelihood criteria are closely related to conventional information criteria such as the Schwarz criteria, but (instead of evaluating them at the maximum likelihood estimate) use the posterior and are based on the expected value of the log of the likelihood function. Like information criteria, such features do not involve the prior (except insofar as the prior enters the posterior and, thus, the MCMC algorithm) and, thus, will be less sensitive to prior choice (and can be considered as approximations to the log of the marginal likelihood).

Finally, we turn to the calculation of impulse responses. In linear (time-invariant) VARs, impulse responses can be taken directly from the Vector Moving Average (VMA) representation implied by the VAR. However, with a TVP-VAR the implied VMA is changing over time. Suppose the VMA representation of a standard VAR is given by:

$$y_t = \sum_{i=0}^{\infty} \theta_i u_{t-i},$$

then the usual result is that an impulse response h periods in the future is the appropriate element of θ_h. With a TVP-VAR the implied VMA will, of course, have time varying coefficients:

$$y_t = \sum_{i=0}^{\infty} \theta_{t-i,i} u_{t-i}.$$

This raises two issues when calculating impulse responses. The first is that the impulse responses will be changing over time. Hence, we have to either plot impulse responses for every time period or choose a few time periods for detailed study. We adopt both these strategies in our empirical work. A second and more subtle issue arises due to the treatment of shocks other than the one being perturbed. To explain this issue, suppose we are interested in the effect of a shock of size one (to the structural errors in the measurement equation) which occurs at time τ on the variables at time $\tau + h$. Strictly speaking, an impulse response is usually interpreted as a difference in conditional expectations such as:

$$E\left(y_{\tau+h}|I_\tau, u_\tau = 1\right) - E\left(y_{\tau+h}|I_\tau\right),$$

where I_τ denotes information through time τ. In any nonlinear time series model, these expectations can be calculated using simulation methods (as in Koop, 1996). However, this can be computationally demanding, so it is much easier to simply take the structural VAR coefficients at time τ (i.e. α_τ and Υ_τ) and calculate a conventional impulse response function. In linear models, these two strategies are identical, but with nonlinear models they can be slightly different. Nevertheless, in this paper we adopt this second simpler strategy. Formally, it can be interpreted as an impulse response function calculated assuming all shocks to the model (including the shocks to the state equations) between time τ and $\tau + h$ are simply set to their expected values of zero.

Empirical Appendix

The tables in this appendix give point estimates (posterior medians) and measures of uncertainty (i.e. the 10th and 90th percentile of each posterior) of the VAR coefficients (α_t) and the lower-triangular Choleski decomposition of the error covariance matrix in the structural VAR (i.e. $\Upsilon_t\Upsilon = A_t^{-1}\Sigma_t$). With regards to the former, note that we are working with two lags and, to keep the tables as brief as possible, do not present results for the intercepts. Following Primiceri (2005), we present results for 1975Q1, 1981Q4 and 1996Q1. Given that we have extended the data set, we also present results for 2006Q3.

To aid in interpretation of the tables, note that any parameter will relate to one of our three equations which we call the inflation (dp), unemployment rate (u) and interest rate (r) equations, respectively. It can also relate to any variable. We use this equation/variable notation to identify parameters in the tables. So, for instance, the $(1, 3)$ element of Υ_t is in the interest rate equation and is the coefficient on the error in the inflation equation. This we label as r/dp. For the VAR coefficients, we will have, e.g., the coefficient on the second lag of the unemployment rate in the inflation equation. This we label as $\mathrm{dp}_t/\mathrm{u}_{t-2}$.

Eq./Var.	VAR	1975Q1	1981Q4	1996Q1	2006Q3
dp_t/u_{t-1}	-0.33 (-0.41,-0.24)	-0.20 (-0.34,-0.07)	-0.22 (-0.34,-0.08)	-0.24 (-0.37,-0.11)	-0.26 (-0.42,-0.11)
dp_t/r_{t-1}	-0.06 (-0.09,-0.03)	0.09 (0.00,0.18)	0.04 (-0.03,0.11)	0.00 (-0.08,0.08)	-0.07 (-0.19,0.06)
dp_t/dp_{t-1}	1.53 (1.47,1.59)	1.26 (1.10,1.41)	1.17 (0.99,1.32)	1.07 (0.89,1.22)	0.86 (0.67,1.04)
u_t/u_{t-1}	1.32 (1.14,1.48)	0.99 (0.82,1.16)	1.04 (0.87,1.21)	0.91 (0.73,1.10)	0.96 (0.96,1.21)
u_t/r_{t-1}	0.00 (-0.06,0.06)	-0.07 (-0.17,0.04)	0.06 (-0.02,0.13)	-0.12 (-0.23,0.01)	-0.13 (-0.31,0.05)
u_t/dp_{t-1}	0.03 (-0.09,0.16)	-0.02 (-0.16,0.13)	-0.11 (-0.29,0.06)	0.04 (-0.14,0.22)	0.06 (-0.15,0.27)
r_t/du_{t-1}	-1.02 (-1.26,-0.77)	-0.45 (-0.68,-0.22)	-0.44 (-0.69,-0.21)	-0.36 (-0.61,-0.12)	-0.50 (-0.79,-0.21)
r_t/r_{t-1}	0.71 (0.63,0.81)	0.80 (0.64,0.97)	0.85 (0.66,1.01)	1.08 (0.87,1.27)	0.96 (0.73,1.22)
r_t/dp_{t-1}	0.57 (0.38,0.75)	0.36 (0.10,0.61)	0.45 (0.19,0.71)	0.27 (0.04,0.51)	0.00 (-0.29,0.28)
dp_t/u_{t-2}	0.27 (0.19,0.35)	0.16 (0.04,0.29)	0.16 (0.03,0.27)	0.18 (0.05,0.30)	0.12 (-0.03,0.27)
dp_t/r_{t-2}	0.06 (0.03,0.09)	0.02 (-0.08,0.11)	-0.01 (-0.07,0.05)	-0.03 (-0.10,0.04)	-0.01 (-0.15,0.11)
dp_t/dp_{t-2}	-0.52 (-0.59,-0.46)	-0.34 (-0.51,-0.17)	-0.27 (-0.42,-0.10)	-0.18 (-0.31,-0.02)	0.02 (-0.16,0.20)
u_t/u_{t-2}	-0.39 (-0.54,-0.22)	-0.07 (-0.22,0.09)	-0.14 (-0.30,0.01)	0.00 (-0.17,0.16)	-0.06 (-0.27,0.17)
u_t/r_{t-2}	0.03 (-0.03,0.09)	0.14 (0.04,0.25)	0.01 (-0.06,0.08)	0.13 (0.01,0.24)	0.15 (-0.02,0.33)
u_t/dp_{t-2}	0.00 (-0.14,0.13)	0.02 (-0.12,0.15)	0.10 (-0.06,0.27)	0.00 (-0.17,0.18)	0.01 (-0.19,0.22)
r_t/u_{t-2}	1.01 (0.77,1.24)	0.33 (0.11,0.18)	0.37 (0.14,0.59)	0.24 (0.01,0.49)	0.24 (-0.05,0.54)
r_t/r_{t-2}	0.20 (0.10,0.28)	0.02 (-0.14,0.18)	0.05 (-0.10,0.22)	-0.16 (-0.34,0.03)	-0.10 (-0.35,0.13)
r_t/dp_{t-2}	-0.44 (-0.63,-0.25)	-0.22 (-0.46,0.04)	-0.25 (-0.49,0.00)	-0.11 (-0.33,0.12)	0.07 (-0.20,0.36)

Table A1: Posterior Median of VAR Coefficients ($10^{th}/90^{th}$ percentiles in parentheses)

Table A2: Posterior Median of Υ_t ($10^{th}/90^{th}$ percentiles in parentheses)					
Eq./Var.	VAR	1975Q1	1981Q4	1996Q1	2006Q3
dp/dp	0.21	0.30	0.42	0.07	0.26
	(0.11,0.39)	(0.07,0.96)	(0.08,0.97)	(0.03,0.17)	(0.11,0.67)
u/dp	-0.01	0.01	-0.10	-0.01	-0.02
	(-0.04,0.02)	(-0.06,0.14)	(-0.34,0.03)	(-0.06,0.01)	(-0.12,0.05)
r/dp	0.05	-0.09	0.88	0.09	-0.03
	(0.01,0.13)	(-0.61.0.20)	(-0.05,2.22)	(-0.02,0.29)	(-0.27,0.15)
u/u	0.42	1.02	0.11	0.12	0.09
	(0.23,0.78)	(0.59,2.00)	(0.03,0.46)	(0.04,0.33)	(0.03,0.39)
r/u	-0.42	-1.03	0.02	-0.03	-0.05
	(-0.80,-0.21)	(-2.24,-0.34)	(-0.34,0.44)	(-0.23,0.09)	(-0.34,0.04)
r/r	0.39	0.15	0.20	0.10	0.08
	(0.21,0.78)	(0.03,1.37)	(0.04,1.88)	(0.03,0.46)	(0.02,0.52)