Bayesian Multivariate Time Series Methods for Empirical Macroeconomics

Koop, Gary and Korobilis, Dimitris

University of Strathclyde

27 September 2009

Bayesian Multivariate Time Series Methods for Empirical Macroeconomics*

Gary Koop
University of Strathclyde

Dimitris Korobilis
University of Strathclyde

September 2009

Abstract

Macroeconomic practitioners frequently work with multivariate time series models such as VARs, factor augmented VARs as well as time-varying parameter versions of these models (including variants with multivariate stochastic volatility). These models have a large number of parameters and, thus, over-parameterization problems may arise. Bayesian methods have become increasingly popular as a way of overcoming these problems. In this monograph, we discuss VARs, factor augmented VARs and time-varying parameter extensions and show how Bayesian inference proceeds. Apart from the simplest of VARs, Bayesian inference requires the use of Markov chain Monte Carlo methods developed for state space models and we describe these algorithms. The focus is on the empirical macroeconomist and we offer advice on how to use these models and methods in practice and include empirical illustrations. A website provides Matlab code for carrying out Bayesian inference in these models.

Keywords: Empirical macroeconometrics, Bayesian estimation, MCMC, vector autoregressions, factor models, time-varying parameters

JEL Classification: C11, C13, C15, C51, C52, C53, C87, E52

1 Introduction

Most questions of interest to empirical macroeconomists involve several variables and, thus, must be addressed using multivariate time series methods. Many different multivariate time series models have been used in macroeconomics, but since the pioneering work of Sims (1980), Vector Autoregressive (VAR) models have been among the most popular. It soon became apparent that, in many applications, the assumption that the VAR coefficients were constant over time might be a poor one. For instance, in practice, it is often found that the macroeconomy of the 1960s and 1970s was different from the 1980s and 1990s. This led to an increased interest in models which allowed for time variation in the VAR coefficients and time-varying parameter VARs (TVP-VARs) arose. In

*Both authors are Fellows at the Rimini Centre for Economic Analysis. Address for correspondence: Gary Koop, Department of Economics, University of Strathclyde, 130 Rottenrow, Glasgow G4 0GE, UK. Email: Gary.Koop@strath.ac.uk

addition, in the 1980s many industrialized economies experienced a reduction in the volatility of many macroeconomic variables. This Great Moderation of the business cycle led to an increasing focus on appropriate modelling of the error covariance matrix in multivariate time series models and this led to the incorporation of multivariate stochastic volatility in many empirical papers. In 2008 many economies went into recession and many of the associated policy discussions suggest that the parameters in VARs may be changing again.

Macroeconomic data sets typically involve monthly, quarterly or annual observations and, thus are only of moderate size. But VARs have a great number of parameters to estimate. This is particularly true if the number of dependent variables is more than two or three (as is required for an appropriate modelling of many macroeconomic relationships). Allowing for time-variation in VAR coefficients causes the number of parameters to proliferate. Allowing for the error covariance matrix to change over time only increases worries about over-parameterization. The research challenge facing macroeconomists is how to build models that are flexible enough to be empirically relevant, capturing key data features such as the Great Moderation, but not so flexible as to be seriously over-parameterized. Many approaches have been suggested, but a common theme in most of these is shrinkage. Whether for forecasting or estimation, it has been found that shrinkage can be of great benefit in reducing over-parameterization problems. This shrinkage can take the form of imposing restrictions on parameters or shrinking them towards zero. This has initiated a large increase in the use of Bayesian methods since prior[1] information provides a logical and formally consistent way of introducing shrinkage. Furthermore, the computational tools necessary to carry out Bayesian estimation of high dimensional multivariate time series models have become well-developed and, thus, models which may have been difficult or impossible to estimate ten or twenty years ago can now be routinely used by macroeconomic practitioners.

A related class of models, and associated worries about over-parameterization, has arisen due to the increase in data availability. Macroeconomists are able to work with hundreds of different time series variables collected by government statistical agencies and other policy institutes. Building a model with hundreds of time series variables (with at most a few hundred observations on each) is a daunting task, raising the issue of a potential proliferation of parameters and a need for shrinkage or other methods for reducing the dimensionality of the model. Factor methods, where the information in the hundreds of variables is distilled into a few factors, are a popular way of dealing with this problem. Combining factor methods with VARs results in Factor-augmented VARs or FAVARs. However, just as with VARs, there is a need to allow for time-variation in parameters, which leads to an interest in TVP-FAVARs. Here, too, Bayesian methods are popular and for the same reason as with TVP-VARs: Bayesian priors provide a sensible way of avoiding over-parameterization problems and Bayesian computational tools are well-designed for dealing with such

[1] Prior information can be purely subjective. However, as will be discussed below, often empirical Bayesian or hierarchical priors are used by macroeconomists. For instance, the state equation in a state space model can be interpreted as a hierarchical prior.

models.

In this monograph, we survey, discuss and extend the Bayesian literature on VARs, TVP-VARs and TVP-FAVARs with a focus on the practitioner. That is, we go beyond simply defining each model, but specify how to use them in practice, discuss the advantages and disadvantages of each and offer some tips on when and why each model can be used. In addition to this, we discuss some new modelling approaches for TVP-VARs. A website contains Matlab code which allows for Bayesian estimation of the models discussed in this monograph. Bayesian inference often involves the use of Markov chain Monte Carlo (MCMC) posterior simulation methods such as the Gibbs sampler. For many of the models, we provide complete details in this monograph. However, in some cases we only provide an outline of the MCMC algorithm. Complete details of all algorithms are given in a manual on the website.

Empirical macroeconomics is a very wide field and VARs, TVP-VARs and factor models, although important, are only some of the tools used in the field. It is worthwhile briefly mentioning what we are not covering in this monograph. There is virtually nothing in this monograph about macroeconomic theory and how it might infuse econometric modelling. For instance, Bayesian estimation of dynamic stochastic general equilibrium (DSGE) models is very popular. There will be no discussion of DSGE models in this monograph (see An and Schorfheide, 2007 or Del Negro and Schorfheide, 2009 for excellent treatments of Bayesian DSGE methods). Also, macroeconomic theory is often used to provide identifying restrictions to turn reduced form VARs into structural VARs suitable for policy analysis. We will not discuss structural VARs, although some of our empirical examples will provide impulse responses from structural VARs using standard identifying assumptions.

There is also a large literature on what might, in general, be called regime-switching models. Examples include Markov switching VARs, threshold VARs, smooth transition VARs, floor and ceiling VARs, etc. These, although important, are not discussed here.

The remainder of this monograph is organized as follows. Section 2 provides discussion of VARs to develop some basic insights into the sorts of shrinkage priors (e.g. the Minnesota prior) and methods of finding empirically-sensible restrictions (e.g. stochastic search variable selection, or SSVS) that are used in empirical macroeconomics. Our goal is to extend these basic methods and priors used with VARs, to TVP variants. However, before considering these extensions, Section 3 discusses Bayesian inference in state space models using MCMC methods. We do this since TVP-VARs (including variants with multivariate stochastic volatility) are state space models and it is important that the practitioner knows the Bayesian tools associated with state space models before proceeding to TVP-VARs. Section 4 discusses Bayesian inference in TVP-VARs, including variants which combine the Minnesota prior or SSVS with the standard TVP-VAR. Section 5 discusses factor methods, beginning with the dynamic factor model, before proceeding to the factor augmented VAR (FAVAR) and TVP-FAVARs. Empirical illustrations are used throughout and Matlab code for implementing these illustrations (or, more generally, doing Bayesian

inference in VARs, TVP-VARs and TVP-FAVARs) is available on the website associated with this monograph.[2]

2 Bayesian VARs

2.1 Introduction and Notation

The VAR(p) model can be written as:

$$y_t = a_0 + \sum_{j=1}^{p} A_j y_{t-j} + \varepsilon_t \tag{1}$$

where y_t for $t = 1, .., T$ is an $M \times 1$ vector containing observations on M time series variables, ε_t is an $M \times 1$ vector of errors, a_0 is an $M \times 1$ vector of intercepts and A_j is an $M \times M$ matrix of coefficients. We assume ε_t to be i.i.d. $N(0, \Sigma)$. Exogenous variables or more deterministic terms (e.g. deterministic trends or seasonals) can easily be added to the VAR and included in all the derivations below, but we do not do so to keep the notation as simple as possible.

The VAR can be written in matrix form in different ways and, depending on how this is done, some of the literature expresses results in terms of the multivariate Normal and others in terms of the matric-variate Normal distribution (see, e.g. Canova, 2007, and Kadiyala and Karlsson, 1997). The former arises if we use an $MT \times 1$ vector y which stacks all T observations on the first dependent variable, then all T observations on the second dependent variable, etc.. The latter arises if we define Y to be a $T \times M$ matrix which stacks the T observations on each dependent variable in columns next to one another. ε and E denote stackings of the errors in a manner conformable to y and Y, respectively. Define $x_t = \left(1, y'_{t-1}, .., y'_{t-p}\right)$ and

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_T \end{bmatrix}. \tag{2}$$

Note that, if we let $K = 1 + Mp$ be the number of coefficients in each equation of the VAR, then X is a $T \times K$ matrix.

Finally, if $A = (a_0 \; A_1 \; .. \; A_p)'$ we define $\alpha = vec(A)$ which is a $KM \times 1$ vector which stacks all the VAR coefficients (and the intercepts) into a vector. With all these definitions, we can write the VAR either as:

$$Y = XA + E \tag{3}$$

or

$$y = (I_M \otimes X)\alpha + \varepsilon, \tag{4}$$

[2] The website address is: http://personal.strath.ac.uk/gary.koop/bayes_matlab_code_by_koop_and_korobilis.html

where $\varepsilon \sim N(0, \Sigma \otimes I_M)$.

The likelihood function can be derived and shown to be of a form that breaks into two parts: one a distribution for α given Σ and another where Σ^{-1} has a Wishart distribution.[3] That is,

$$\alpha|\Sigma, y \sim N\left(\widehat{\alpha}, \Sigma \otimes (X'X)^{-1}\right) \tag{5}$$

and

$$\Sigma^{-1}|y \sim W\left(S^{-1}, T-K-M-1\right), \tag{6}$$

where $\widehat{A} = (X'X)^{-1} X'Y$ is the OLS estimate of A and $\widehat{\alpha} = vec\left(\widehat{A}\right)$ and

$$S = \left(Y - X\widehat{A}\right)' \left(Y - X\widehat{A}\right).$$

2.2 Priors

A variety of priors can be used with the VAR, of which we discuss some useful ones below. They differ in relation to three issues.

First, VARs are not parsimonious models. They have a great many coefficients. For instance, α contains KM parameters which, for a VAR(4) involving 5 dependent variables is 105. With quarterly macroeconomic data, the number of observations on each variable might be at most a few hundred. Without prior information, it is hard to obtain precise estimates of so many coefficients and, thus, features such as impulse responses and forecasts will tend to be imprecisely estimated (i.e. posterior or predictive standard deviations can be large). For this reason, it can be desirable to "shrink" forecasts and prior information offers a sensible way of doing this shrinkage. The priors discussed below differ in the way they achieve this goal.

Second, the priors used with VARs differ in whether they lead to analytical results for the posterior and predictive densities or whether MCMC methods are required to carry out Bayesian inference. With the VAR, natural conjugate priors lead to analytical results, which can greatly reduce the computational burden. Particularly if one is carrying out a recursive forecasting exercise which requires repeated calculation of posterior and predictive distributions, non-conjugate priors which require MCMC methods can be very computationally demanding.

Third, the priors differ in how easily they can handle departures from the unrestricted VAR given in (1) such as allowing for different equations to have different explanatory variables, allowing for VAR coefficients to change over time, allowing for heteroskedastic structures for the errors of various sorts, etc. Natural conjugate priors typically do not lend themselves to such extensions.

[3] In this monograph, we use standard notational conventions to define all distributions such as the Wishart. See, among many other places, the appendix to Koop, Poirier and Tobias (2007). Wikipedia is also a quick and easy source of information about distributions.

2.2.1 The Minnesota Prior

Early work with Bayesian VARs with shrinkage priors was done by researchers at the University of Minnesota or the Federal Reserve Bank of Minneapolis (see Doan, Litterman and Sims (1984) and Litterman (1986)). The priors they used have come to be known as Minnesota priors. They are based on an approximation which leads to great simplifications in prior elicitation and computation. This approximation involves replacing Σ with an estimate, $\widehat{\Sigma}$. The original Minnesota prior simplifies even further by assuming Σ to be a diagonal matrix. In this case, each equation of the VAR can be estimated one at a time and we can set $\widehat{\sigma}_{ii} = s_i^2$ (where s_i^2 is the standard OLS estimate of the error variance in the i^{th} equation and $\widehat{\sigma}_{ii}$ is the ii^{th} element of $\widehat{\Sigma}$). When Σ is not assumed to be diagonal, a simple estimate such as $\widehat{\Sigma} = \frac{S}{T}$ can be used.

When Σ is replaced by an estimate, we only have to worry about a prior for α and the Minnesota prior assumes:

$$\alpha \sim N\left(\underline{\alpha}_{Min}, \underline{V}_{Min}\right). \tag{7}$$

The Minnesota prior can be thought of as a way of automatically choosing $\underline{\alpha}_{Min}$ and $\underline{V}_{Min}$ in a manner which is sensible in many empirical contexts. To explain the Minnesota prior, note first that the explanatory variables in the VAR in any equation can be divided into the own lags of the dependent variable, the lags of the other dependent variables and exogenous or deterministic variables (in 1 the intercept is the only exogenous or deterministic variable, but in general there can be more such variables).

For the prior mean, $\underline{\alpha}_{Min}$, the Minnesota prior involves setting most or all of its elements to zero (thus ensuring shrinkage of the VAR coefficients towards zero and lessening the risk of over-fitting). When using growth rates data (e.g. GDP growth, the growth in the money supply, etc., which are typically found to be stationary and exhibit little persistence), it is sensible to simply set $\underline{\alpha}_{Min} = 0_{KM}$. However, when using levels data (e.g. GDP, the money supply, etc.) the Minnesota prior uses a prior mean expressing a belief that the individual variables exhibit random walk behavior. Thus, $\underline{\alpha}_{Min} = 0_{KM}$ except for the elements corresponding to the first own lag of the dependent variable in each equation. These elements are set to one. These are the traditional choices for $\underline{\alpha}_{Min}$, but anything is possible. For instance, in our empirical illustration we set the prior mean for the coefficient on the first own lag to be 0.9, reflecting a prior belief that our variables exhibit a fair degree of persistence, but not unit root behavior.

The Minnesota prior assumes the prior covariance matrix, $\underline{V}_{Min}$, to be diagonal. If we let $\underline{V}_i$ denote the block of $\underline{V}_{Min}$ associated with the K coefficients in equation i and $\underline{V}_{i,jj}$ be its diagonal elements, then a common implementation of the Minnesota prior would set:

$$\underline{V}_{i,jj} = \begin{cases} \frac{a_1}{p^2} \text{ for coefficients on own lags} \\ \frac{a_2 \sigma_{ii}}{p^2 \sigma_{jj}} \text{ for coefficients on lags of variable } j \neq i \\ a_3 \sigma_{ii} \text{ for coefficients on exogenous variables} \end{cases}. \tag{8}$$

This prior simplifies the complicated choice of fully specifying all the elements of $\underline{V}_{Min}$ to choosing three scalars, $\underline{a}_1, \underline{a}_2, \underline{a}_3$. This form captures the sensible properties that, as lag length increases, coefficients are increasingly shrunk towards zero and that (by setting $\underline{a}_1 > \underline{a}_2$) own lags are more likely to be important predictors than lags of other variables. The exact choice of values for $\underline{a}_1, \underline{a}_2, \underline{a}_3$ depends on the empirical application at hand and the researcher may wish to experiment with different values for them. Typically, the researcher sets $\sigma_{ii} = s_i^2$. Litterman (1986) provides much additional motivation and discussion of these choices (e.g. an explanation for how the term $\frac{\sigma_{ii}}{\sigma_{jj}}$ adjusts for differences in the units that the variables are measured in).

Many variants of the Minnesota prior have been used in practice (e.g. Kadiyala and Karlsson (1997) divide prior variances by p instead of the p^2 which is used in (8)) as researchers make slight adjustments to tailor the prior for their particular application. The Minnesota prior has enjoyed a recent boom in popularity because of its simplicity and success in many applications, particularly involving forecasting. For instance, Banbura, Giannone and Reichlin (2008) use a slight modification of the Minnesota prior in a large VAR with over 100 dependent variables. Typically, factor methods are used with such large panels of data, but Banbura et al (2008) find that the Minnesota prior leads to even better forecasting performance than factor methods.

A big advantage of the Minnesota prior is that it leads to simple posterior inference involving only the Normal distribution. It can be show that the posterior for α has the form:

$$\alpha|y \sim N\left(\overline{\alpha}_{Min}, \overline{V}_{Min}\right) \tag{9}$$

where

$$\overline{V}_{Min} = \left[\underline{V}_{Min}^{-1} + \left(\widehat{\Sigma}^{-1} \otimes (X'X)\right)\right]^{-1}$$

and

$$\overline{\alpha}_{Min} = \overline{V}_{Min}\left[\underline{V}_{Min}^{-1}\underline{\alpha}_{Min} + \left(\widehat{\Sigma}^{-1} \otimes X\right)' y\right].$$

But a disadvantage of the Minnesota prior is that it does not provide a full Bayesian treatment of Σ as an unknown parameter. Instead it simply plugs in $\Sigma = \widehat{\Sigma}$, ignoring any uncertainty in this parameter. In the remainder of this section we will discuss methods which treat Σ as an unknown parameter. However, as we shall see, this (apart from one restrictive special case) means that analytical methods are not available and MCMC methods are required).

2.2.2 Natural conjugate priors

Natural conjugate priors are those where the prior, likelihood and posterior come from the same family of distributions. Our previous discussion of the

likelihood function (see equations 5 and 6) suggests that, for the VAR, the natural conjugate prior has the form:

$$\alpha|\Sigma \sim N\left(\underline{\alpha}, \Sigma \otimes \underline{V}\right) \tag{10}$$

and

$$\Sigma^{-1} \sim W\left(\underline{S}^{-1}, \underline{\nu}\right) \tag{11}$$

where $\underline{\alpha}, \underline{V}, \underline{\nu}$ and $\underline{S}$ are prior hyperparameters chosen by the researcher.

With this prior the posterior becomes:

$$\alpha|\Sigma, y \sim N\left(\overline{\alpha}, \Sigma \otimes \overline{V}\right) \tag{12}$$

and

$$\Sigma^{-1}|y \sim W\left(\overline{S}^{-1}, \overline{\nu}\right) \tag{13}$$

where

$$\overline{V} = \left[\underline{V}^{-1} + X'X\right]^{-1},$$

$$\overline{A} = \overline{V}\left[\underline{V}^{-1}\underline{A} + X'X\widehat{A}\right],$$

$\overline{\alpha} = vec\left(\overline{A}\right)$,

$$\overline{S} = S + \underline{S} + \widehat{A}'X'X\widehat{A} + \underline{A}'\underline{V}^{-1}\underline{A} - \overline{A}'\left(\underline{V}^{-1} + X'X\right)\overline{A}$$

and

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

In the previous formulae, we use notation where $\underline{A}$ is a $K \times M$ matrix made by unstacking the $KM \times 1$ vector $\underline{\alpha}$.

Posterior inference about the VAR coefficients can be carried out using the fact that the marginal posterior (i.e. after integrating out Σ) for α is a multivariate t-distribution. The mean of this t-distribution is $\overline{\alpha}$, its degrees of freedom parameter is $\overline{\nu}$ and its covariance matrix is:

$$var\left(\alpha|y\right) = \frac{1}{\overline{\nu} - M - 1}\overline{S} \otimes \overline{V}.$$

These facts can be used to carry out carry out posterior inference in this model.

The predictive distribution for y_{T+1} in this model has an analytical form and, in particular, is multivariate-t with $\overline{\nu}$ degrees of freedom. The predictive mean of y_{T+1} is $\left(x_{T+1}\overline{A}\right)'$ which can be used to produce point forecasts. The predictive covariance matrix is $\frac{1}{\overline{\nu}-2}\left[1 + x_{T+1}\overline{V}x'_{T+1}\right]\overline{S}$. When forecasting more than one period ahead, an analytical formula for the predictive density does not exist. This means that either the direct forecasting method must be used (which turns the problem into one which only involves one step ahead forecasting) or predictive simulation is required.

Any values for the prior hyperparameters, $\underline{\alpha}, \underline{V}, \underline{\nu}$ and $\underline{S}$, can be chosen. The noninformative prior is obtained by setting $\underline{\nu} = \underline{S} = \underline{V}^{-1} = cI$ and letting $c \rightarrow 0$. It can be seen that this leads to posterior and predictive results which are based on familiar OLS quantities. The drawback of the noninformative prior is that it does not do any of the shrinkage which we have argued is so important for VAR modelling.

Thus, for the natural conjugate prior, analytical results exist which allow for Bayesian estimation and prediction. There is no need to use posterior simulation algorithms unless interest centers on nonlinear functions of the parameters (e.g. impulse response analysis such as those which arise in structural VARs, see Koop, 1992). The posterior distribution of, e.g., impulse responses can be obtained by Monte Carlo integration. That is, draws of Σ^{-1} can be obtained from (13) and, conditional on these, draws of α can be taken from (12).[4] Then draws of impulse responses can be calculated using these drawn values of Σ^{-1} and α.

However, there are two properties of this prior that can be undesirable in some circumstances. The first is that the $(I_M \otimes X)$ form of the explanatory variables in (4) means that every equation must have the same set of explanatory variables. For an unrestricted VAR this is fine, but is not appropriate if the researcher wishes to impose restrictions. Suppose, for instance, the researcher is working with a VAR involving variables such as output growth and the growth in the money supply and wants to impose a strict form of the neutrality of money. This would imply that the coefficients on the lagged money growth variables in the output growth equation are zero (but coefficients of lagged money growth in other equations would not be zero). Such restrictions cannot be imposed with the natural conjugate prior described here.

To explain the second possibly undesirable property of this prior, we introduce notation where individual elements of Σ are denoted by σ_{ij}. The fact that the prior covariance matrix has the form $\Sigma \otimes \underline{V}$ (which is necessary to ensure natural conjugacy of the prior), implies that the prior covariance of the coefficients in equation i is $\sigma_{ii}\underline{V}$. This means that the prior covariance of the coefficients in any two equations must be proportional to one another, a possibly restrictive feature. In our example, the researcher believing in the neutrality of money may wish to proceed as follows: in the output growth equation, the prior mean of the coefficients on lagged money growth variables should be zero and the prior covariance matrix should be very small (i.e. expressing a prior belief that these coefficients are very close to zero). In other equations, the prior covariance matrix on the coefficients on lagged money growth should be much larger. The natural conjugate prior does not allow us to use prior information of this form. It also does not allow us to use the Minnesota prior. That is, the Minnesota prior covariance matrix in (8) is written in terms of blocks which were labelled $\underline{V}_{i,jj}$ involving i subscripts. That is, these blocks vary across equations which is not allowed for in the natural conjugate prior.

[4] Alternatively, draws of α can be directly taken from its multivariate-t marginal posterior distribution.

These two properties should be kept in mind when using the natural conjugate prior. There are generalizations of this natural conjugate prior, such as the extended natural conjugate prior of Kadiyala and Karlsson (1997), which surmount these problems. However, these lose the huge advantage of the natural conjugate prior described in this section: that analytical results are available and so no posterior simulation is required.

A property of natural conjugate priors is that, since the prior and likelihood have the same distributional form, the prior can be considered as arising from a fictitious sample. For instance, a comparison of (5) and (10) shows that $\widehat{\alpha}$ and $(X'X)^{-1}$ in the likelihood play the same role as $\underline{\alpha}$ and $\underline{V}$ in the prior. The latter can be interpreted as arising from a fictitious sample (also called "dummy observations"), Y_0 and X_0 (e.g. $\underline{V} = (X_0'X_0)^{-1}$ and $\underline{\alpha}$ based on an OLS estimate $(X_0'X_0)^{-1} X_0'Y_0$). This interpretation is developed in papers such as Sims (1993) and Sims and Zha (1998). On one level, this insight can simply serve as another way of motivating choices for $\underline{\alpha}$ and $\underline{V}$ as arising from particular choices for Y_0 and X_0. But papers such as Sims and Zha (1998) show how the dummy observation approach can be used to elicit priors for structural VARs. In this monograph, we will focus on the econometric as opposed to the macroeconomic issues. Accordingly, we will work with reduced form VARs and not say much about structural VARs. Here we only note that posterior inference in structural VARs is usually based on a reduced form VAR such as that discussed here, but then coefficients are transformed so as to give them a structural interpretation (see, e.g., Koop, 1992, for a simple example). For instance, structural VARs are often written as:

$$C_0 y_t = c_0 + \sum_{j=1}^{p} C_j y_{t-j} + u_t \tag{14}$$

where u_t is i.i.d. $N(0, I)$. Given appropriate identifying restrictions, there will be a mapping from the parameters of the reduced form VAR in (1) to the structural VAR. Thus, Bayesian inference can be done by using posterior simulation methods in the reduced form VAR and transforming each draw into a draw from the structural VAR.

While discussing such macroeconomic issues, it is worth noting that there is a growing literature that uses the insights of economic theory (e.g. from real business cycle or DSGE models) to elicit priors for VARs. Prominent examples include Ingram and Whiteman (1994) and Del Negro and Schorfheide (2004). We will not discuss this work in this monograph.

Finally, it is also worth mentioning the work of Villani (2009) on steady state priors for VARs. We have motivated prior information as being important as a way of ensuring shrinkage in an over-parameterized VAR. However, most of the shrinkage discussed previously relates to the VAR coefficients. Often researchers have strong prior information about the unconditional means (i.e. the steady states) of the variables in their VARs. It is desirable to include such information as an additional source of shrinkage in the VAR. However, it is not easy to do this in the VAR in (1) since the intercepts cannot be directly interpreted as the

unconditional means of the variables in the VAR. Villani (2009) recommends writing the VAR as:

$$\widetilde{A}(L)(y_t - \widetilde{a}_0) = \varepsilon_t \tag{15}$$

where $\widetilde{A}(L) = I - \widetilde{A}_1 L - .. - \widetilde{A}_p L^p$, L is the lag operator and ε_t is i.i.d. $N(0, \Sigma)$. In this parameterization, $\widetilde{a}_0$ can be interpreted as the vector of unconditional means of the dependent variables and a prior placed on it reflecting the researcher's beliefs about steady state values for them. For $\widetilde{A}(L)$ and Σ one of the priors described previously (or below) can be used. A drawback of this approach is that an analytical form for the posterior no longer exists. However, Villani (2009) develops a Gibbs sampling algorithm for carrying out Bayesian inference in this model.

2.2.3 The Independent Normal-Wishart Prior

The natural conjugate prior has the large advantage that analytical results are available for posterior inference and prediction. However, it does have the drawbacks noted previously (i.e. it assumes each equation to have the same explanatory variables and it restricts the prior covariance of the coefficients in any two equations to be proportional to one another). Accordingly, in this section, we introduce a more general framework for VAR modelling. To carry out Bayesian inference in these models will require posterior simulation algorithms such as the Gibbs sampler. The natural conjugate prior had $\alpha|\Sigma$ being Normal and Σ^{-1} being Wishart. Note that the fact that the prior for α depends on Σ implies that α and Σ are not independent of one another. In this section, we work with a prior which has VAR coefficients and the error covariance being independent of one another (hence the name "independent Normal-Wishart prior").

To allow for different equations in the VAR to have different explanatory variables, we have to modify our previous notation slightly. To avoid any possibility of confusion, we will use "β" as notation for VAR coefficients in this restricted VAR model instead of α. We write each equation of the VAR as:

$$y_{mt} = z'_{mt}\beta_m + \varepsilon_{mt},$$

with $t = 1, .., T$ observations for $m = 1, .., M$ variables. y_{mt} is the t^{th} observation on the m^{th} variable, z_{mt} is a k_m-vector containing the t^{th} observation of the vector of explanatory variables relevant for the m^{th} variable and β_m is the accompanying k_m-vector of regression coefficients. Note that if we had $z_{mt} = \left(1, y'_{t-1}, .., y'_{t-p}\right)'$ for $m = 1, .., M$ then we would obtain the unrestricted VAR of the previous section. However, by allowing for z_{mt} to vary across equations we are allowing for the possibility of a restricted VAR (i.e. it allows for some of the coefficients on the lagged dependent variables to be restricted to zero).

We can stack all equations into vectors/matrices as $y_t = (y_{1t}, .., y_{Mt})'$, $\varepsilon_t = (\varepsilon_{1t}, .., \varepsilon_{Mt})'$,

$$\beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_M \end{pmatrix},$$

$$Z_t = \begin{pmatrix} z'_{1t} & 0 & \cdots & 0 \\ 0 & z'_{2t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & z'_{Mt} \end{pmatrix},$$

where β is a $k \times 1$ vector and Z_t is $M \times k$ where $k = \sum_{j=1}^{M} k_j$. As before, we assume ε_t to be i.i.d. $N(0, \Sigma)$.

Using this notation, we can write the (possibly restricted) VAR as:

$$y_t = Z_t \beta + \varepsilon_t. \tag{16}$$

Stacking as:

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix},$$

$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix},$$

$$Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_T \end{pmatrix}$$

we can write

$$y = Z\beta + \varepsilon$$

and ε is $N(0, I \otimes \Sigma)$.

It can be seen that the restricted VAR can be written as a Normal linear regression model with an error covariance matrix of a particular form. A very general prior for this model (which does not involve the restrictions inherent in the natural conjugate prior) is the independent Normal-Wishart prior:

$$p\left(\beta, \Sigma^{-1}\right) = p(\beta)\, p\left(\Sigma^{-1}\right)$$

where

$$\beta \sim N\left(\underline{\beta}, \underline{V}_\beta\right) \tag{17}$$

and

$$\Sigma^{-1} \sim W\left(\underline{S}^{-1}, \underline{\nu}\right). \tag{18}$$

Note that this prior allows for the prior covariance matrix, $\underline{V}_\beta$, to be anything the researcher chooses, rather than the restrictive $\Sigma \otimes \underline{V}$ form of the natural conjugate prior. For instance, the researcher could set $\underline{\beta}$ and $\underline{V}_\beta$ exactly as in the Minnesota prior. A noninformative prior can be obtained by setting $\underline{\nu} = \underline{S} = \underline{V}_\beta^{-1} = 0$.

Using this prior, the joint posterior $p\left(\beta, \Sigma^{-1}|y\right)$ does not have a convenient form that would allow easy Bayesian analysis (e.g. posterior means and variances do not have analytical forms). However, the conditional posterior distributions $p\left(\beta|y, \Sigma^{-1}\right)$ and $p\left(\Sigma^{-1}|y, \beta\right)$ do have convenient forms:

$$\beta|y, \Sigma^{-1} \sim N\left(\overline{\beta}, \overline{V}_\beta\right), \tag{19}$$

where

$$\overline{V}_\beta = \left(\underline{V}_\beta^{-1} + \sum_{t=1}^{T} Z_t' \Sigma^{-1} Z_t\right)^{-1}$$

and

$$\overline{\beta} = \overline{V}_\beta \left(\underline{V}_\beta^{-1} \underline{\beta} + \sum_{i=1}^{T} Z_t' \Sigma^{-1} y_t\right).$$

Furthermore,

$$\Sigma^{-1}|y, \beta \sim W\left(\overline{S}^{-1}, \overline{\nu},\right) \tag{20}$$

where

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

and

$$\overline{S} = \underline{S} + \sum_{t=1}^{T} \left(y_t - Z_t\beta\right)\left(y_t - Z_t\beta\right)'.$$

Accordingly, a Gibbs sampler which sequentially draws from the Normal $p\left(\beta|y, \Sigma\right)$ and the Wishart $p\left(\Sigma^{-1}|y, \beta\right)$ can be programmed up in a straightforward fashion. As with any Gibbs sampler, the resulting posterior simulator output can be used to calculate posterior properties of any function of the parameters, marginal likelihoods (for model comparison) and/or to do prediction.

Note that, for the VAR, Z_τ will contain lags of variables and, thus, contain information dated $\tau - 1$ or earlier. The one-step ahead predictive density (i.e. the one for predicting at time τ given information through $\tau - 1$), conditional on the parameters of the model is:

$$y_\tau | Z_\tau, \beta, \Sigma \sim N\left(Z_t\beta, \Sigma\right).$$

This result, along with a Gibbs sampler producing draws $\beta^{(r)}, \Sigma^{(r)}$ for $r = 1, .., R$ allows for predictive inference.[5] For instance, the predictive mean (a popular point forecast) could be obtained as:

$$E\left(y_\tau | Z_\tau\right) = \frac{\sum_{r=1}^{R} Z_t\beta^{(r)}}{R}$$

and other predictive moments can be calculated in a similar fashion. Alternatively, predictive simulation can be done at each Gibbs sampler draw, but this can be computationally demanding. For forecast horizons greater than one, the direct method can be used. This strategy for doing predictive analysis can be used with any of the priors or models discussed below.

2.2.4 Stochastic Search Variable Selection (SSVS) in VARs

SSVS as Implemented in George, Sun and Ni (2008) In the previous sections, we have described various priors for unrestricted and restricted VARs which allow for shrinkage of VAR coefficients. However, these approaches required substantial prior input from the researcher (although this prior input can be of an automatic form such as in the Minnesota prior). There is another prior that, in a sense, does shrinkage and leads to restricted VARs, but does so in an automatic fashion that requires only minimal prior input from the researcher. The methods associated with this prior are called SSVS and are enjoying increasing popularity and, accordingly, we describe them here in detail. SSVS can be done in several ways. Here we describe the implementation of George, Sun and Ni (2008).

The basic idea underlying SSVS can be explained quite simply. Suppose α_j is a VAR coefficient. Instead of simply using a prior for it as before (e.g. as in (10), SSVS specifies a hierarchical prior (i.e. a prior expressed in terms of parameters which in turn have a prior of their own) which is a mixture of two Normal distributions:

$$\alpha_j | \gamma_j \sim \left(1 - \gamma_j\right) N\left(0, \kappa_{0j}^2\right) + \gamma_j N\left(0, \kappa_{1j}^2\right), \tag{21}$$

where γ_j is a dummy variable. If γ_j equals one then α_j is drawn from the second Normal and if it equals zero then α_j is drawn from the first Normal. The prior is hierarchical since γ_j is treated as an unknown parameter and estimated in a data-based fashion. The SSVS aspect of this prior arises by choosing the first prior variance, κ_{0j}^2, to be "small" (so that the coefficient is constrained to be virtually zero) and the second prior variance, κ_{1j}^2, to be "large" (implying a relatively noninformative prior for the corresponding coefficient). Below we describe what George, Sun and Ni (2008) call a "default semi-automatic approach" to

[5] Typically, some initial draws are discarded as the "burn in". Accordingly, $r = 1, .., R$ should be the post-burn in draws.

choosing κ_{0j}^2 and κ_{1j}^2 which requires minimal subjective prior information from the researcher.

The SSVS approach can be thought of as automatically selecting a restricted VAR since it can, in a data-based fashion, set $\gamma_j = 0$ and (to all intents and purposes) delete the corresponding lagged dependent variable form the model. Alternatively, SSVS can be thought of as a way of doing shrinkage since VAR coefficients can be shrunk to zero.

SSVS can be used to select a single restricted model (e.g. the researcher can select a restricted VAR which contains only those lagged dependent variables whose coefficients have $\Pr\left(\gamma_j = 1|y\right) > a$ for some choice of a such as $a = 0.5$). Alternatively, if the MCMC algorithm described below is simply run and posterior results for the VAR coefficients calculated using the resulting MCMC output, the result will be Bayesian model averaging (BMA).

SSVS allows us to work with the unrestricted VAR and have the algorithm pick out an appropriate restricted VAR. Accordingly we will return to our notation for the unrestricted VAR (see Section 2.1). The unrestricted VAR is written in (3) and α is the $KM \times 1$ vector of VAR coefficients. SSVS can be interpreted as defining a hierarchical prior for all of the elements of α and Σ. The prior for α given in (21) can be written more compactly as:

$$\alpha|\gamma \sim N\left(0, DD\right), \tag{22}$$

where γ is a $KM \times 1$ vector of unknown parameters with typical element $\gamma_j \in \{0, 1\}$, and D is a diagonal matrix with $(j,j)^{th}$ element given by d_j where

$$d_j = \begin{cases} \kappa_{0j} \text{ if } \gamma_j = 0 \\ \kappa_{1j} \text{ if } \gamma_j = 1 \end{cases}. \tag{23}$$

Note that this prior implies a mixture of two Normals as written in (21).

George, Sun and Ni (2008) describe a "default semi-automatic approach" to selecting the prior hyperparameters κ_{0j} and κ_{1j} which involves setting $\kappa_{0j} = c_0\sqrt{\widehat{var}(\alpha_j)}$ and $\kappa_{1j} = c_1\sqrt{\widehat{var}(\alpha_j)}$ where $\widehat{var}(\alpha_j)$ is an estimate of the variance of the coefficient in an unrestricted VAR (e.g. the ordinary least squares quantity or an estimate based on a preliminary Bayesian estimation the VAR using a noninformative prior). The pre-selected constants c_0 and c_1 must have $c_0 \ll c_1$ (e.g. $c_0 = 0.1$ and $c_1 = 10$).

For $\gamma = (\gamma_1, .., \gamma_{KM})'$, the SSVS prior assumes that each element has a Bernoulli form (independent of the other elements of γ) and, hence, for $j = 1, .., KM$, we have

$$\begin{array}{l} \Pr\left(\gamma_j = 1\right) = \underline{q}_j \\ \Pr\left(\gamma_j = 0\right) = 1 - \underline{q}_j \end{array}. \tag{24}$$

A natural default choice is $\underline{q}_j = 0.5$ for all j, implying each coefficient is *a priori* equally likely to be included as excluded.

So far, we have said nothing about the prior for Σ and (for the sake of brevity) we will not provide details relating to it. Suffice it to note here that if a

Wishart prior for Σ^{-1} like (18) is used, then a formula very similar to (20) can be used as a block in a Gibbs sampling algorithm. Alternatively, George, Sun and Ni (2008) use a prior for Σ which allows for them to do SSVS on the error covariance matrix. That is, although they always assume the diagonal elements of Σ are positive (so as to ensure a positive definite error covariance matrix), they allow for parameters which determine the off-diagonal elements to have an SSVS prior thus allowing for restrictions to be imposed on Σ. We refer the interested reader to George, Sun and Ni (2008) or the manual on the website associated with this monograph for details.

Posterior computation in the VAR with SSVS prior can be carried out using a Gibbs sampling algorithm. For the VAR coefficients we have

$$\alpha|y,\gamma,\Sigma \sim N(\overline{\alpha}_\alpha, \overline{V}_\alpha), \tag{25}$$

where

$$\overline{V}_\alpha = [\Sigma^{-1} \otimes (X'X) + (DD)^{-1}]^{-1},$$

$$\overline{\alpha}_\alpha = \overline{V}_\alpha[(\Psi\Psi') \otimes (X'X)\hat{\alpha}],$$

$$\hat{A} = (X'X)^{-1}X'Y$$

and

$$\hat{\alpha} = vec(\hat{A}).$$

The conditional posterior for γ has γ_j being independent Bernoulli random variables:

$$\begin{aligned} \Pr\left[\gamma_j = 1|y,\alpha\right] &= \overline{q}_j, \\ \Pr\left[\gamma_j = 0|y,\alpha\right] &= 1 - \overline{q}_j, \end{aligned} \tag{26}$$

where

$$\overline{q}_j = \frac{\frac{1}{\kappa_{1j}} \exp\left(-\frac{\alpha_j^2}{2\kappa_{1j}^2}\right) \underline{q}_j}{\frac{1}{\kappa_{1j}} \exp\left(-\frac{\alpha_j^2}{2\kappa_{1j}^2}\right) \underline{q}_j + \frac{1}{\kappa_{0j}} \exp\left(-\frac{\alpha_j^2}{2\kappa_{0j}^2}\right) \left(1 - \underline{q}_j\right)}.$$

Thus, a Gibbs sampler involving the Normal distribution and the Bernoulli distribution (and the Wishart distribution if a Wishart prior is used for Σ^{-1}) allows for posterior inference in this model.

SSVS as Implemented in Korobilis (2009b) The implementation of SSVS just described is a popular one. However, there are other similar methods for automatic model selection in VARs. In particular, the approach of George, Sun and Ni (2008) involves selecting values for the "small" prior variance κ_{0j}. The reader may ask why not set "small" exactly equal to zero? This has been done in regression models in papers such as Kuo and Mallick (1997) through restricting coefficients to be precisely zero if $\gamma_j = 0$. There are some subtle statistical issues which arise when doing this.[6] Korobilis (2009b) has extended the use of such methods to VARs. Since, unlike the implementation of George, Sun and Ni (2008), this approach leads to restricted VARs (as opposed to unrestricted VARs with very tight priors on some of the VAR coefficients), we return to our notation for restricted VARs and modify it slightly. In particular, replace (16) by

$$y_t = Z_t\widetilde{\beta} + \varepsilon_t. \tag{27}$$

where $\widetilde{\beta} = \widetilde{D}\beta$ and $\widetilde{D}$ is a diagonal matrix with the j^{th} diagonal element being γ_j (where, as before, γ_j is a dummy variable). In words, this model allows for each VAR coefficient to be set to zero (if $\gamma_j = 0$) or included in an unrestricted fashion (if $\gamma_j = 1$).

Bayesian inference using the prior can be carried out in a straightforward fashion. For exact details on the necessary MCMC algorithm, see Korobilis (2009b) and the manual on the website associated with this book. However, the idea underlying this algorithm can be explained quite simply. Conditional on γ, this model is a restricted VAR and the MCMC algorithm of Section 2.2.2 for the independent Normal-Wishart prior can be used. Thus, all that is required is a method for taking draws from γ (conditional on the parameters of the VAR). Korobilis (2009b) derives the necessary distribution.

2.3 Empirical Illustration of Bayesian VAR Methods

To illustrate Bayesian VAR methods using some of the priors and methods described above, we use a quarterly US data set on the inflation rate $\Delta\pi_t$ (the annual percentage change in a chain-weighted GDP price index), the unemployment rate u_t (seasonally adjusted civilian unemployment rate, all workers over age 16) and the interest rate r_t (yield on the three month Treasury bill rate). Thus $y_t = (\Delta\pi_t, u_t, r_t)'$. The sample runs from 1953Q1 to 2006Q3. These three variables are commonly used in New Keynesian VARs.[7] Examples of papers which use these, or similar, variables include Cogley and Sargent (2005), Primiceri (2005) and Koop, Leon-Gonzalaz and Strachan (2009). The data are plotted in Figure 1.

[6] For instance, asympotically such priors will always set $\gamma_j = 1$ for all j.

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

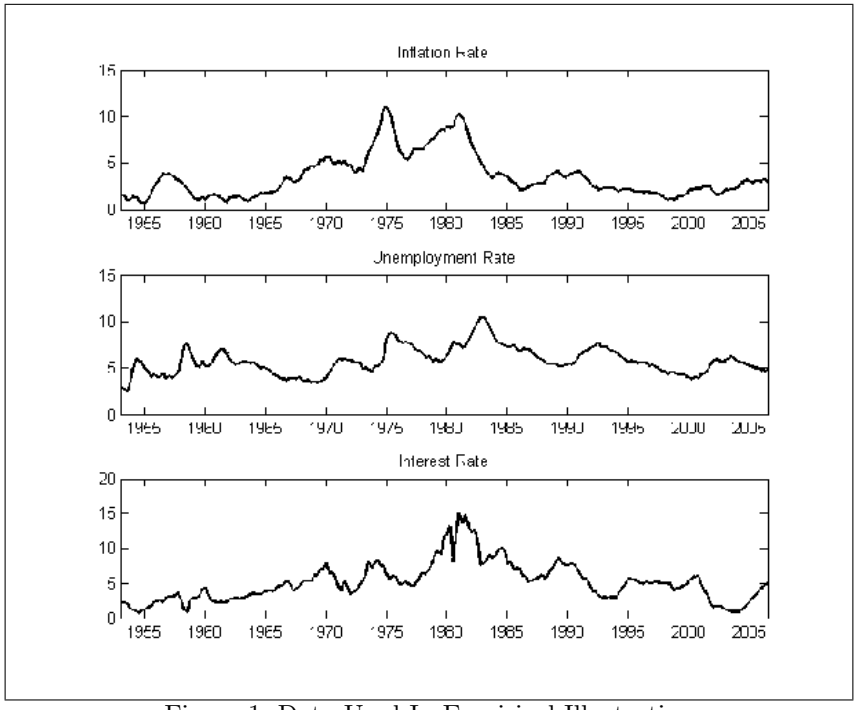

Figure 1: Data Used In Empirical Illustration

To illustrate Bayesian VAR analysis using this data, we work with an unrestricted VAR with an intercept and four lags of all variables included in every equation and consider the following six priors:

- Noninformative: Noninformative version of natural conjugate prior (equations 10 and 11 with $\underline{\alpha} = 0_{KM\times 1}$, $\underline{V} = 0_{K\times K}$, $\underline{v} = 0$ and $\underline{S} = 0_{M\times M}$).
- Natural conjugate: Informative natural conjugate prior with subjectively chosen prior hyperparameters (equations 10 and 11 with $\underline{\alpha} = 0_{KM\times 1}$, $\underline{V} = 10I_K$, $\underline{v} = M + 1$ and $\underline{S}^{-1} = I_M$).
- Minnesota: Minnesota prior (equations 7 and 8, where $\underline{\alpha}_{Min}$ is zero, except for the first own lag of each variable which is 0.9. Σ is diagonal with elements s_i^2 obtained from univariate regressions of each dependent variable on an intercept and four lags of all variables).
- Independent Normal-Wishart: Independent Normal-Wishart prior with subjectively chosen prior hyperparameters (equations 17 and 18 with $\underline{\beta} = 0_{KM\times 1}$, $\underline{V}_\beta = 10I_{KM}$, $\underline{v} = M + 1$ and $\underline{S}^{-1} = I_M$).
- SSVS-VAR: SSVS prior for VAR coefficients (with default semi-automatic approach prior with $c_0 = 0.1$ and $c_1 = 10$) and Wishart prior for Σ^{-1} (equation 18 with $\underline{v} = M + 1$ and $\underline{S}^{-1} = I_M$).

- SSVS: SSVS on both VAR coefficients and error covariance (default semi-automatic approach).[8]

For the first three priors, analytical posterior and predictive results are available. For the last three, posterior and predictive simulation is required. The results below are based on 50000 MCMC draws, for which the first 20000 are discarded as burn-in draws. For impulse responses (which are nonlinear functions of the VAR coefficients and Σ), posterior simulation methods are required for all six priors.

With regards to impulse responses, they are identified by assuming C_0 in (14) is lower triangular and the dependent variables are ordered as: inflation, unemployment and interest rate. This is a standard identifying assumption used, among many others, by Bernanke and Mihov (1998), Christiano, Eichanbaum and Evans (1999) and Primiceri (2005). It allows for the interpretation of the interest rate shock as a monetary policy shock.

With VARs, the parameters themselves (as opposed to functions of them such as impulse responses) are rarely of direct interest. In addition, the fact that there are so many of them makes it hard for the reader to interpret tables of VAR coefficients. Nevertheless, Table 1 presents posterior means of all the VAR coefficients for two priors: the noninformative one and SSVS prior. Note that they are yielding similar results, although there is some evidence that SSVS is slightly shrinking the coefficients towards zero.

Table 1. Posterior mean of VAR Coefficients for Two Priors

	Noninformative			SSVS - VAR		
	$\Delta\pi_t$	u_t	r_t	$\Delta\pi_t$	u_t	r_t
Intercept	0.2920	0.3222	-0.0138	0.2053	0.3168	0.0143
$\Delta\pi_{t-1}$	1.5087	0.0040	0.5493	1.5041	0.0044	0.3950
u_{t-1}	-0.2664	1.2727	-0.7192	-0.142	1.2564	-0.5648
r_{t-1}	-0.0570	-0.0211	0.7746	-0.0009	-0.0092	0.7859
$\Delta\pi_{t-2}$	-0.4678	0.1005	-0.7745	-0.5051	0.0064	-0.226
u_{t-2}	0.1967	-0.3102	0.7883	0.0739	-0.3251	0.5368
r_{t-2}	0.0626	-0.0229	-0.0288	0.0017	-0.0075	-0.0004
$\Delta\pi_{t-3}$	-0.0774	-0.1879	0.8170	-0.0074	0.0047	0.0017
u_{t-3}	-0.0142	-0.1293	-0.3547	0.0229	-0.0443	-0.0076
r_{t-3}	-0.0073	0.0967	0.0996	-0.0002	0.0562	0.1119
$\Delta\pi_{t-4}$	0.0369	0.1150	-0.4851	-0.0005	0.0028	-0.0575
u_{t-4}	0.0372	0.0669	0.3108	0.0160	0.0140	0.0563
r_{t-4}	-0.0013	-0.0254	0.0591	-0.0011	-0.0030	0.0007

Remember that SSVS allows to the calculation of $\Pr\left(\gamma_j = 1|y\right)$ for each VAR coefficient and such posterior inclusion probabilities can be used either in model averaging or in model selection. Table 2 presents such posterior inclusion probabilities using the SSVS-VAR prior. The empirical researcher may wish to

[8] SSVS on the non-diagonal elements of Σ is not fully described in this monograph. See George, Sun and Ni (2008) for complete details.

present such a table for various reasons. For instance, if the researcher wishes to select a single restricted VAR which only includes coefficients with $\Pr\left(\gamma_j = 1|y\right) > \frac{1}{2}$, then he would work with a model which restricts 25 of 39 coefficients to zero. Table 2 shows which coefficients are important. Of the 14 included coefficients two are intercepts and three are first own lags in each equation. The researcher using SSVS to select a single model would restrict most of the remaining VAR coefficients to be zero. The researcher using SSVS to do model averaging would, in effect, be restricting them to be approximately zero. Note also that SSVS can be used to do lag length selection in an automatic fashion. None of the coefficients on the fourth lag variables is found to be important and only one of nine possible coefficients on third lags is found to be important.

Table 2. Posterior Inclusion Probabilities for VAR Coefficients: SSVS-VAR Prior

	$\Delta\pi_t$	u_t	r_t
Intercept	0.7262	0.9674	0.1029
$\Delta\pi_{t-1}$	1	0.0651	0.9532
u_{t-1}	0.7928	1	0.8746
r_{t-1}	0.0612	0.2392	1
$\Delta\pi_{t-2}$	0.9936	0.0344	0.5129
u_{t-2}	0.4288	0.9049	0.7808
r_{t-2}	0.0580	0.2061	0.1038
$\Delta\pi_{t-3}$	0.0806	0.0296	0.1284
u_{t-3}	0.2230	0.2159	0.1024
r_{t-3}	0.0416	0.8586	0.6619
$\Delta\pi_{t-4}$	0.0645	0.0507	0.2783
u_{t-4}	0.2125	0.1412	0.2370
r_{t-4}	0.0556	0.1724	0.1097

With VARs, the researcher is often interested in forecasting. It is worth mentioning that often recursive forecasting exercises, which involve forecasting at time $\tau = \tau_0, .., T$, are often done. These typically involve estimating a model $T - \tau_0$ times using appropriate sub-samples of the data. If MCMC methods are required, this can be computationally demanding. That is, running an MCMC algorithm $T - \tau_0$ times can (depending on the model and application) be very slow. If this is the case, then the researcher may be tempted to work with methods which do not require MCMC such as the Minnesota or natural conjugate priors. Alternatively, sequential importance sampling methods such as the particle filter (see, e.g. Doucet, Godsill and Andrieu, 2000 or Johannes and Polson, 2009) can be used which do not require the MCMC algorithm to be run at each point in time.

Table 3 presents predictive results for an out-of-sample forecasting exercise based on the predictive density $p\left(y_{T+1}|y_1.., y_T\right)$ where $T = 2006Q3$. It can be seen that for this empirical example, which involves a moderately large data set, the prior is having relatively little impact. That is, predictive means and standard deviations are similar for all six priors, although it can be seen that the

predictive standard deviations with the Minnesota prior do tend to be slightly smaller than the other priors.

Table 3. Predictive mean of y_{T+1} (st. dev. in parentheses)			
PRIOR	$\Delta\pi_{T+1}$	u_{T+1}	r_{T+1}
Noninformative	3.105 (0.315)	4.610 (0.318)	4.382 (0.776)
Minnesota	3.124 (0.302)	4.628 (0.319)	4.350 (0.741)
Natural conjugate	3.106 (0.313)	4.611 (0.314)	4.380 (0.748)
Indep. Normal-Wishart	3.110 (0.322)	4.622 (0.324)	4.315 (0.780)
SSVS - VAR	3.097 (0.323)	4.641 (0.323)	4.281 (0.787)
SSVS	3.108 (0.304)	4.639 (0.317)	4.278 (0.785)
True value, y_{T+1}	3.275	4.700	4.600

Figures 2 and 3 present impulse responses of all three of our variables to all three of the shocks for two of the priors: the noninformative one and the SSVS prior. In these figures the posterior median is the solid line and the dotted lines are the 10^{th} and 90^{th} percentiles. These impulse responses all have sensible shapes, similar to those found by other authors. The two priors are giving similar results, but a careful examination of them do reveal some differences. Especially at longer horizons, there is evidence that SSVS leads to slightly more precise inferences (evidenced by a narrower band between the 10^{th} and 90^{th} percentiles) due to the shrinkage it provides.

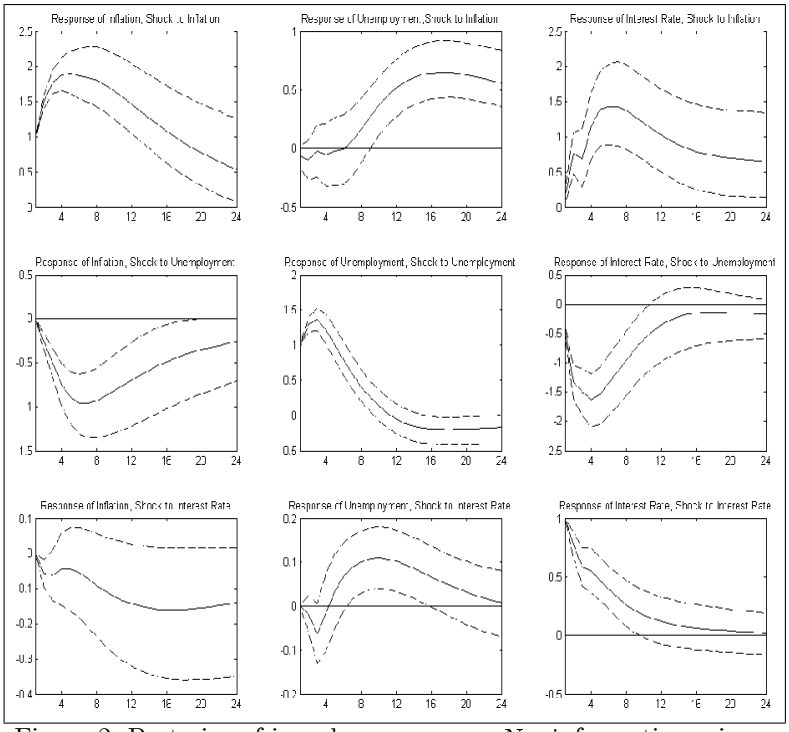

Figure 2: Posterior of impulse responses - Noninformative prior.

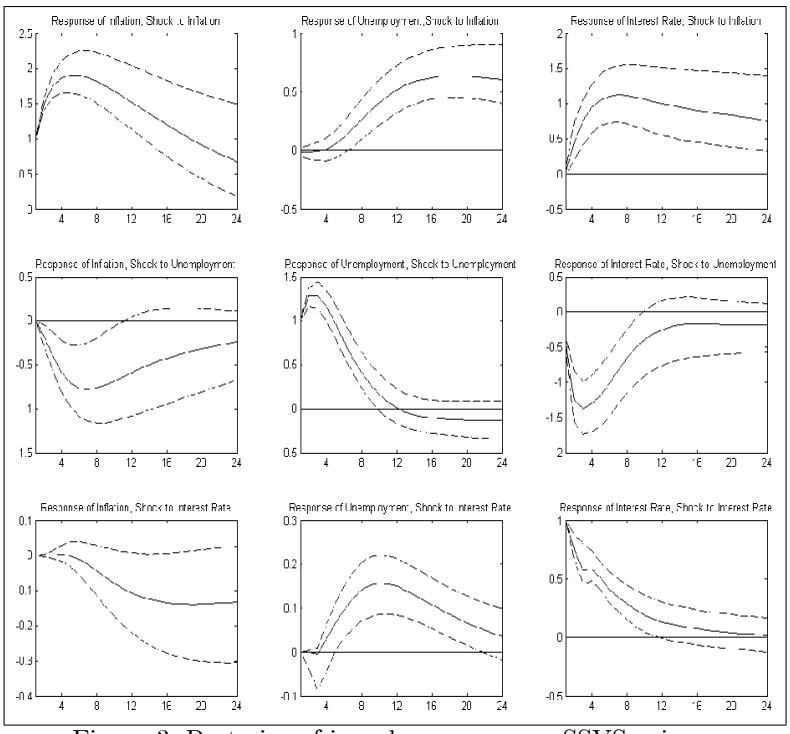

Figure 3: Posterior of impulse responses - SSVS prior.

3 Bayesian State Space Modeling and Stochastic Volatility

3.1 Introduction and Notation

In the section on Bayesian VAR modeling, we showed that the (possibly restricted) VAR could be written as:

$$y_t = Z_t\beta + \varepsilon$$

for appropriate definitions of Z_t and β. In many macroeconomic applications, it is undesirable to assume β to be constant, but it is sensible to assume that β evolves gradually over time. A standard version of the TVP-VAR which will be discussed in the next section extends the VAR to:

$$y_t = Z_t\beta_t + \varepsilon_t,$$

where

$$\beta_{t+1} = \beta_t + u_t.$$

Thus, the VAR coefficients are allowed to vary gradually over time. This is a state space model.

Furthermore, previously we assumed ε_t to be i.i.d. $N(0, \Sigma)$ and, thus, the model was homoskedastic. In empirical macroeconomics, it is often important to allow for the error covariance matrix to change over time (e.g. due to the Great Moderation of the business cycle) and, in such cases, it is desirable to assume ε_t to be i.i.d. $N(0, \Sigma_t)$ so as to allow for heteroskedasticity. This raises the issue of stochastic volatility which, as we shall see, also leads us into the world of state space models.

These considerations provide a motivation for why we must provide a section on state space models before proceeding to TVP-VARs and other models of more direct relevance for empirical macroeconomics. We begin this section by first discussing Bayesian methods for the Normal linear state space model. These methods can be used to model evolution of the VAR coefficients in the TVP-VAR. Unfortunately, stochastic volatility cannot be written in the form of a Normal linear state space model. Thus, after briefly discussing nonlinear state space modelling in general, we present Bayesian methods for particular nonlinear state space models of interest involving stochastic volatility.

We will adopt a notational convention commonly used in the state space literature where, if a_t is a time t quantity (i.e. a vector of states or data) then $a^t = (a_1', .., a_t')'$ stacks all the a_ts up to time t. So, for instance, y^T will denote the entire sample of data on the dependent variables and β^T the vector containing all the states.

3.2 The Normal Linear State Space Model

A general formulation for the Normal linear state space model (which contains the TVP-VAR defined above as a special case) is:

$$y_t = W_t\delta + Z_t\beta_t + \varepsilon_t, \tag{28}$$

and

$$\beta_{t+1} = T_t\beta_t + u_t, \tag{29}$$

where y_t is an $M \times 1$ vector containing observations on M time series variables, ε_t is an $M \times 1$ vector of errors, W_t is a known $M \times p_0$ matrix (e.g. this could contain lagged dependent variables or other explanatory variables with constant coefficients), δ is a $p_0 \times 1$ vector of parameters. Z_t is a known $M \times k$ matrix (e.g. this could contain lagged dependent variables or other explanatory variables with time varying coefficients), β_t is a $k \times 1$ vector of parameters which evolve over time (these are known as states). We assume ε_t to be independent $N(0, \Sigma_t)$ and u_t to be a $k \times 1$ vector which is independent $N(0, Q_t)$. ε_t and u_s are independent of one another for all s and t. T_t is a $k \times k$ matrix which is typically treated as known, but occasionally T_t is treated as a matrix of unknown parameters.

Equations (28) and (29) define a state space model. Equation (28) is called the measurement equation and (29) the state equation. Models such as this have been used for a wide variety of purposes in econometrics and many other fields. The interested reader is referred to West and Harrison (1997) and Kim and Nelson (1999) for a broader Bayesian treatment of state space models than that provided here. Harvey (1989) and Durbin and Koopman (2001) provide good non-Bayesian treatments of state space models.

For our purpose, the important thing to note is that, for given values of δ, T_t, Σ_t and Q_t (for $t = 1, .., T$), various algorithms have been developed which allow for posterior simulation of β_t for $t = 1, .., T$. Popular and efficient algorithms are described in Carter and Kohn (1994), Fruhwirth-Schnatter (1994), DeJong and Shephard (1995) and Durbin and Koopman (2002). Since these are standard and well-understood algorithms, we will not present complete details here. In the Matlab code on the website associated with this monograph, the algorithm of Carter and Kohn (1994) is used. These algorithms can be used as a block in an MCMC algorithm to provide draws from the posterior of β_t conditional on δ, T_t, Σ_t and Q_t (for $t = 1, .., T$). The exact treatment of δ, T_t, Σ_t and Q_t depends on the empirical application at hand. The standard TVP-VAR fixes some of these to known values (e.g. $\delta = 0, T_t = I$ are common choices) and treats others as unknown parameters (although it usually restricts $Q_t = Q$ and, in the case of the homoskedastic TVP-VAR additionally restricts $\Sigma_t = \Sigma$ for all t). An MCMC algorithm is completed by taking draws of the unknown parameters from their posteriors (conditional on the states). The next part of this section elaborates on how this MCMC algorithm works. To focus on state space model issues, the algorithm is for the case where δ is a vector of unknown parameters, $Q_t = Q$ and $\Sigma_t = \Sigma$ and T_t is known.

An examination of (28) reveals that, if β_t for $t = 1, .., T$ were known (as opposed to being unobserved), then the state space model would reduce to a multivariate Normal linear regression model:

$$y_t^* = W_t\delta + \varepsilon_t,$$

where $y_t^* = y_t - Z_t\beta_t$. Thus, standard results for the multivariate Normal linear regression model could be used, except the dependent variable would be y_t^* instead of y_t. This suggests that an MCMC algorithm can be set up for the state space model. That is, $p\left(\delta|y^T, \Sigma, \beta^T\right)$ and $p\left(\Sigma^{-1}|y^T, \delta, \beta^T\right)$ will typically have a simple textbook form. Below we will use the independent Normal-Wishart prior for δ and Σ^{-1}. This was introduced in our earlier discussion of VAR models.

Note next that a similar reasoning can be used for the covariance matrix for the error in the state equation. That is, if β_t for $t = 1, .., T$ were known, then the state equation, (29), is a simple variant of multivariate Normal regression model. This line of reasoning suggests that $p\left(Q^{-1}|y^T, \delta, \beta^T\right)$ will have a simple and familiar form.[9]

[9] The case where T_t contains unknown parameters would involve drawing from

Combining these results for $p\left(\delta|y^T,\Sigma,\beta^T\right)$, $p\left(\Sigma^{-1}|y^T,\delta,\beta^T\right)$ and $p\left(Q^{-1}|y^T,\delta,\beta^T\right)$ with one of the standard methods (e.g. that of Carter and Kohn, 1994) for taking random draws from $p\left(\beta^T|y^T,\delta,\Sigma,Q\right)$ will completely specify an MCMC algorithm which allows for Bayesian inference in the state space model. In the following material we develop such an MCMC algorithm for a particular prior choice, but we stress that other priors can be used with minor modifications.

Here we will use an independent Normal-Wishart prior for δ and Σ^{-1} and a Wishart prior for Q^{-1}. It is worth noting that the state equation can be interpreted as already providing us with a prior for β^T. That is, (29) implies:

$$\beta_{t+1}|\beta_t,Q\sim N\left(T_t\beta_t,Q\right). \tag{30}$$

Formally, the state equation implies the prior for the states is:

$$p\left(\beta^T|Q\right)=\prod_{t=1}^{T}p\left(\beta_t|\beta_{t-1},Q\right)$$

where the terms on the right-hand side are given by (30). This is an example of a hierarchical prior, since the prior for β^T depends on the Q which, in turn, requires its own prior.

One minor issue should be mentioned: that of initial conditions. The prior for β_1 depends on β_0. There are standard ways of treating this issue. For instance, if we assume $\beta_0=0$, then the prior for β_1 becomes:

$$\beta_1|Q\sim N\left(0,Q\right).$$

Similarly, authors such as Carter and Kohn (1994) simply assume β_0 has some unspecified distribution as its prior. Alternatively, in the TVP-VAR (or any TVP regression model) we can simply set $\beta_1=0$ and $W_t=Z_t$.[10]

Combining these prior assumptions together, we have

$$p\left(\delta,\Sigma,Q,\beta^T\right)=p\left(\delta\right)p\left(\Sigma\right)p\left(Q\right)p\left(\beta^T|Q\right)$$

where

$$\delta\sim N\left(\underline{\delta},\underline{V}\right), \tag{31}$$

$$\Sigma^{-1}\sim W\left(\underline{S}^{-1},\underline{\nu}\right), \tag{32}$$

and

$$Q^{-1}\sim W\left(\underline{Q}^{-1},\underline{\nu}_Q\right). \tag{33}$$

$p\left(Q,T_1,..,T_T|y,\beta_1,..,\beta_T\right)$ which can usually be done fairly easily. In the time-invariant case where $T_1=..=T_T\equiv\Upsilon$, $p\left(\Upsilon,Q|y,\beta_1,..,\beta_T\right)$ has a from of the same structure as a VAR.

[10] This result follows from the fact that $y_t=Z_t\beta_t+\varepsilon_t$ with β_1 left unrestricted and $y_t=Z_t\delta+Z_t\beta_t+\varepsilon_t$ with $\beta_1=0$ are equivalent models.

The reasoning above suggests that our end goal is an MCMC algorithm which sequentially draws from $p\left(\delta|y^T,\Sigma,\beta^T\right), p\left(\Sigma^{-1}|y^T,\delta,\beta^T\right),\ p\left(Q^{-1}|y^T,\delta,\beta^T\right)$ and $p\left(\beta^T|y^T,\delta,\Sigma,Q\right)$. The first three of these posterior conditional distributions can be dealt with by using results for the multivariate Normal linear regression model. In particular,

$$\delta|y^T,\Sigma,\beta^T \sim N\left(\overline{\delta},\overline{V}\right).$$

where

$$\overline{V}=\left(\underline{V}^{-1}+\sum_{t=1}^{T}W_t'\Sigma^{-1}W_t\right)^{-1}$$

and

$$\overline{\delta}=\overline{V}\left(\underline{V}^{-1}\underline{\delta}+\sum_{t=1}^{T}W_t'\Sigma^{-1}\left(y_t-Z_t\beta_t\right)\right).$$

Next we have

$$\Sigma^{-1}|y^T,\delta,\beta^T \sim W\left(\overline{S}^{-1},\overline{\nu}\right),$$

where

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

and

$$\overline{S}=\underline{S}+\sum_{t=1}^{T}\left(y_t-W_t\delta-Z_t\beta_t\right)\left(y_t-W_t\delta-Z_t\beta_t\right)'.$$

Next,

$$Q^{-1}|y^T,\delta,\beta^T \sim W\left(\overline{Q}^{-1},\overline{\nu}_Q\right)$$

where

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

and

$$\overline{Q}=\underline{Q}+\sum_{t=1}^{T}\left(\beta_{t+1}-T_t\beta_t\right)\left(\beta_{t+1}-T_t\beta_t\right)'.$$

To complete our MCMC algorithm, we need a means of drawing from $p\left(\beta^T|y^T,\delta,\Sigma,Q\right)$. But, as discussed previously, there are several standard algorithms that can be used for doing this. Accordingly, Bayesian inference in the Normal linear state space model can be done in a straightforward fashion. We will draw on these results when we return to the TVP-VAR in a succeeding section of this monograph.

3.3 Nonlinear State Space Models

The Normal linear state space model discussed previously is used by empirical macroeconomists not only when working with TVP-VARs, but also for many other purposes. For instance, Bayesian analysis of dynamic stochastic general equilibrium (DSGE) models has become increasingly popular (see, e.g., An and Schorfheide, 2007 or Fernandes-Villaverde, 2009). Estimation of linearized DSGE models involves working with the Normal linear state space model and, thus, the methods discussed above can be used. However, linearizing of DSGE models is done through first order approximations and, very recently, macroeconomists have expressed an interest in using second order approximations. When this is done the state space model becomes nonlinear (in the sense that the measurement equation has y_t being a nonlinear function of the states). This is just one example of how nonlinear state space models can arise in macroeconomics. There are an increasing number of tools which allow for Bayesian computation in nonlinear state space models (e.g. the particle filter is enjoying increasing popularity see, e.g., Johannes and Polson, 2009). Given the focus of this monograph on TVP-VARs and related models, we will not offer a general discussion of Bayesian methods for nonlinear state space models (see Del Negro and Schorfheide, 2009, and Giordani, Kohn and Pitt, 2009 for further discussion). Instead we will focus on an area of particular interest for the TVP-VAR modeler: stochastic volatility.

Broadly speaking, issues relating to the volatility of errors have obtained an increasing prominence in macroeconomics. This is due partially to the empirical regularities that are often referred to as the Great Moderation of the business cycle (i.e. that the volatilities of many macroeconomic variables dropped in the early 1980s and remained low until recently). But it is also partly due to the fact that many issues of macroeconomic policy hinge on error variances. For instance, the debate on why the Great Moderation occurred is often framed in terms of "good policy" versus "good luck" stories which involve proper modeling of error variances. For these reasons, volatility is important so we will spend some time describing Bayesian methods for handling it.

3.3.1 Univariate Stochastic Volatility

We begin with a discussion of stochastic volatility when y_t is a scalar. Although TVP-VARs are multivariate in nature and, thus, Bayesian methods for multivariate stochastic volatility are required, these use methods for univariate stochastic volatility as building blocks. Accordingly, a Bayesian treatment of univariate stochastic volatility is a useful starting point. In order to focus the discussion, we will assume there are no explanatory variables and, hence, adopt a simple univariate stochastic volatility model[11] which can be written as:

[11] In this section we describe a method developed in Kim, Shephard and Chib (1998) which has become more popular than the pioneering approach of Jacquier, Polson and Rossi (1994). Bayesian methods for extensions of this standard stochastic volatility model (e.g. involving non-Normal errors or leverage effects) can be found in Chib, Nardari and Shephard (2002) and Omori, Chib, Shephard and Nakajima (2007).

$$y_t = \exp\left(\frac{h_t}{2}\right)\varepsilon_t \tag{34}$$

and

$$h_{t+1} = \mu + \phi\left(h_t - \mu\right) + \eta_t, \tag{35}$$

where ε_t is i.i.d. $N\left(0,1\right)$ and η_t is i.i.d. $N\left(0,\sigma_\eta^2\right)$. ε_t and η_s are independent of one another for all s and t.

Note that (34) and (35) is a state space model similar to (28) and (29) where h_t for $t = 1, .., t$ can be interpreted as states. However, in contrast to (28), (34) is not a linear function of the states and, hence, our results for Normal linear state space models cannot be directly used.

Note that this parameterization is such that h_t is the log of the variance of y_t. Since variances must be positive, in order to sensibly have Normal errors in the state equation (35), we must define the state equation as holding for log-volatilities. Note also that μ is the unconditional mean of h_t.

With regards to initial conditions, it is common to restrict the log-volatility process to be stationary and impose $|\phi| < 1$. Under this assumption, it is sensible to have:

$$h_0 \sim N\left(\mu, \frac{\sigma_\eta^2}{1-\phi^2}\right) \tag{36}$$

and the algorithm of Kim, Shephard and Chib (1998) described below uses this specification. However, in the TVP-VAR literature it is common to have VAR coefficients evolving according to random walks and, by analogy, TVP-VAR papers such as Primiceri (2005) often work with (multivariate extensions of) random walk specifications for the log-volatilities and set $\phi = 1$. This simplifies the model since, not only do parameters akin to ϕ not have to be estimated, but also μ drops out of the model. However, when $\phi = 1$, the treatment of the initial condition given in (36) cannot be used. In this case, a prior such as $h_0 \sim N\left(\underline{h}, \underline{V}_h\right)$ is typically used. This requires the researcher to choose $\underline{h}$ and $\underline{V}_h$. This can be done subjectively or, as in Primiceri (2005), an initial "training sample" of the data can be set aside to calibrate values for the prior hyperparameters.

In the development of an MCMC algorithm for the stochastic volatility model, the key part is working out how to draw the states. That is (in a similar fashion as for the parameters in the Normal linear state space model), $p\left(\phi|y^T, \mu, \sigma_\eta^2, h^T\right)$, $p\left(\mu|y^T, \phi, \sigma_\eta^2, h^T\right)$ and $p\left(\sigma_\eta^2|y^T, \mu, \phi, h^T\right)$ have standard forms derived using textbook results for the Normal linear regression model and will not be presented here (see, e.g., Kim, Shephard and Chib, 1998 for exact formulae). To complete an MCMC algorithm, all that we require is a method for taking draws from $p\left(h^T|y^T, \mu, \phi, \sigma_\eta^2\right)$. Kim, Shephard and Chib (1998) provide an efficient method for doing this. To explain the basic ideas underlying this algorithm, note that if we square both sides of the measurement equation, (34), and then take logs we obtain:

$$y_t^* = h_t + \varepsilon_t^*, \tag{37}$$

where[12] $y_t^* = \ln\left(y_t^2\right)$ and $\varepsilon_t^* = \ln\left(\varepsilon_t^2\right)$. Equations (37) and (35) define a state space model which is linear in the states. The only thing which prevents us from immediately using our previous results for the Normal linear state space model is the fact that ε_t^* is not Normal. However, as well shall see, it can be approximated by a mixture of different Normal distributions and this allows us to exploit our earlier results.

Mixtures of Normal distributions are very flexible and have been used widely in many fields to approximate unknown or inconvenient distributions. In the case of stochastic volatility, Kim, Shephard and Chib (1998) show that the distribution of ε_t^*, $p\left(\varepsilon_t^*\right)$ can be well-approximated by:

$$p\left(\varepsilon_t^*\right) \approx \sum_{i=1}^{7} q_i f_N\left(\varepsilon_t^* | m_i, v_i^2\right), \tag{38}$$

where $f_N\left(\varepsilon_t^* | m_i, v_i^2\right)$ is the p.d.f. of a $N\left(m_i, v_i^2\right)$ random variable.[13] Crucially, since ε_t is $N(0,1)$ it follows that ε_t^* involves no unknown parameters and neither does this approximation. Thus, q_i, m_i, v_i^2 for $i = 1, .., 7$ are not parameters to be estimated, but simply numbers given in Table 4 of Kim, Shephard and Chib (1998).

An equivalent way of writing (38) is to introduce component indicator variables, $s_t \in \{1, 2, .., 7\}$ for each element in the Normal mixture and writing:

$$\begin{array}{l} \varepsilon_t^* | s_t = i \sim N\left(m_i, v_i^2\right) \\ \Pr\left(s_t = i\right) = q_i \end{array},$$

for $i = 1, .., 7$. This formulation provides insight into how the algorithm works. In particular, the MCMC algorithm does not simply draw the log-volatilities from $p\left(h^T | y^T, \mu, \phi, \sigma_\eta^2\right)$, but rather draws them from $p\left(h^T | y^T, \mu, \phi, \sigma_\eta^2, s^T\right)$. This may seem awkward, but has the huge benefit that standard results from the Normal linear state space models such as those described previously in this section can be used. That is, conditional on knowing $s_1, .., s_T$, the algorithm knows which of the seven Normals ε_t^* comes from at each $t = 1, .., T$ and the model becomes a Normal linear state space model. To complete the MCMC algorithm requires a method for drawing from $p\left(s^T | y^T, \mu, \phi, \sigma_\eta^2, h^T\right)$ but this is simple to do since s_t is a discrete distribution with seven points of support. Precise details are given in Kim, Shephard and Chib (1998).

A Digression: Marginal Likelihood Calculation in State Space Models
Marginal likelihoods are the most popular tool for Bayesian model comparison

[12] In practice, it is common to set $y_t^* = \ln\left(y_t^2 + c\right)$ where c is known as an off-set constant set to a small number (e.g. $c = 0.001$) to avoid numerical problems associated with times where y_t^2 is zero or nearly so.

[13] Omori, Chib, Shephard and Nakajima (2007) recommend an even more accurate approximation using a mixture of 10 Normal distributions.

(e.g. Bayes factors are ratios of marginal likelihoods). In this monograph, we focus on estimation and prediction as opposed to model comparison or hypothesis testing. This is partly because state space models such as TVP-VARs are very flexible and can approximate a wide range of data features. Thus, many researchers prefer to treat them as being similar in spirit to nonparametric models: capable of letting the data speak and uncovering an appropriate model (as opposed to working with several parsimonious models and using statistical methods to select a single one). Furthermore, a Bayesian rule of thumb is that the choice of prior matters much less for estimation and prediction than it does for marginal likelihoods. This is particularly true for high-dimensional models such as TVP-VARs where marginal likelihoods can be sensitive to the choice of prior. For this reason, many Bayesians avoid the use of marginal likelihoods in high dimensional models. Even those who wish to do model comparison often use other metrics (e.g. Geweke and Keane, 2007, uses cross-validation).

However, for the researcher who wishes to use marginal likelihoods, note that there are many methods for calculating them that involve the evaluation of the likelihood function at a point. For instance, information criteria are often approximations to marginal likelihoods and these involve calculating the maximum likelihood estimator. The popular methods of marginal likelihood calculation of Chib (1995) and Gelfand and Dey (1994) involve evaluating the likelihood function. In state space models, a question arises as to what likelihood function should be used. In terms of our notation for the Normal linear state space model, $p\left(y^T|\delta,\Sigma,Q,\beta^T\right)$ and $p\left(y^T|\delta,\Sigma,Q\right)$ can both be used as "likelihoods" and either could be used in any of the methods of marginal likelihood calculation just cited.[14] However, using $p\left(y^T|\delta,\Sigma,Q,\beta^T\right)$ to define the likelihood function could potentially lead to very inefficient computation since the parameter space is of such high dimension.[15] Thus, it is desirable to use $p\left(y^T|\delta,\Sigma,Q\right)$ to define the likelihood function. Fortunately, for the Normal linear state space model a formula for $p\left(y^T|\delta,\Sigma,Q\right)$ is available which can be found in textbooks such as Harvey (1989) or Durbin and Koopman (2001).

For the stochastic volatility model, for the same reasons either $p\left(y^T|\phi,\mu,\sigma_\eta^2,h^T\right)$ or $p\left(y^T|\phi,\mu,\sigma_\eta^2\right)$ could be used to define the likelihood function. It is desirable to to use $p\left(y^T|\phi,\mu,\sigma_\eta^2\right)$ but, unfortunately, an analytical expression for it does not exist. Several methods have been used to surmount this problem, but some of them can be quite complicated (e.g. involving using particle filtering methods to integrate out h^T). Berg, Meyer and Yu (2004) discuss these issues in detail and recommend a simple approximation called the Deviance Information Criterion.

It is also worth noting that the MCMC algorithm for the stochastic volatility model is an example of an auxiliary mixture sampler. That is, it introduces an auxiliary set of states, s^T, which results in a mixture of Normals representation.

[14] Fruhwirth-Schnatter and Wagner (2008) refer to the former of these as the *complete data likelihood* and the latter as the *integrated likelihood*.

[15] The non-Bayesian seeking to find the maximum of this likelihood function would also often run into troubles optimizing in such a high dimensional parameter space.

Conditional on these states, the model is a Normal linear state space model. Fruhwirth-Schnatter and Wagner (2008) exploit this Normality (conditional on the auxiliary states) result to develop methods for calculating marginal likelihoods using auxiliary mixture samplers and such methods can be used with stochastic volatility models.

3.3.2 Multivariate Stochastic Volatility

Let us now return to the state space model of (28) and (29) where y_t is an $M \times 1$ vector and ε_t is i.i.d. $N(0, \Sigma_t)$. As we have stressed previously, in empirical macroeconomics it is often very important to allow for Σ_t to be time varying. There are many ways of doing this. Note that Σ_t is an $M \times M$ positive definite matrix with $\frac{M(M+1)}{2}$ distinct elements. Thus, the complete set of Σ_t for $t = 1, .., T$ contains $\frac{TM(M+1)}{2}$ unknown parameters which is a huge number. In one sense, the literature on multivariate stochastic volatility can be thought of as mitigating this proliferation of parameters problems through parametric restrictions and/or priors and working in parameterizations which ensure that Σ_t is always positive definite. Discussions of various approaches can be found in Asai, McAleer and Yu (2006), Chib, Nardari and Shephard (2006) and Chib, Omori and Asai (2009) and the reader is referred to these papers for complete treatments. In this section, we will describe two approaches popular in macroeconomics. The first was popularized by Cogley and Sargent (2005), the second by Primiceri (2005).

To focus on the issues relating to multivariate stochastic volatility, we will consider the model:

$$y_t = \varepsilon_t \tag{39}$$

and ε_t is i.i.d. $N(0, \Sigma_t)$. Before discussing the specifications used by Cogley and Sargent (2005) and Primiceri (2005) for Σ_t we begin with a very simple specification such that

$$\Sigma_t = D_t$$

where D_t is a diagonal matrix with each diagonal element having a univariate stochastic volatility specification. That is, if d_{it} is the i^{th} diagonal element of D_t for $i = 1, .., M$, then we write $d_{it} = \exp(h_{it})$ and

$$h_{i,t+1} = \mu_i + \phi_i (h_{it} - \mu_i) + \eta_{it}, \tag{40}$$

where $\eta_t = (\eta'_{1t}, .., \eta'_{Mt})'$ is i.i.d. $N(0, D_\eta)$ where D_η is a diagonal matrix (so the errors in the state equation are independent of one another). This model is simple to work with in that it simply says that each error follows its own univariate stochastic volatility model, independent of all the other errors. Thus, the Kim, Shephard and Chib (1998) MCMC algorithm can be used one equation at a time.

This model is typically unsuitable for empirical macroeconomic research since it is not appropriate to assume Σ_t to be diagonal. Many interesting

macroeconomic features (e.g. impulse responses) depend on error covariances so assuming them to be zero may be misleading. Some researchers such as Cogley and Sargent (2005) allow for non-zero covariances in a simple way by writing:

$$\Sigma_t = L^{-1} D_t L^{-1\prime} \tag{41}$$

where D_t is a diagonal matrix with diagonal elements being the error variances and L is a lower triangular matrix with ones on the diagonal. For instance, in the $M = 3$ case we have

$$L = \begin{bmatrix} 1 & 0 & 0 \\ L_{21} & 1 & 0 \\ L_{31} & L_{32} & 1 \end{bmatrix}.$$

This form is particularly attractive for computation since, even through ε_{it} and ε_{jt} (which are the i^{th} and j^{th} elements of ε_t) are no longer independent of one another, we can transform (39) as

$$Ly_t = L\varepsilon_t \tag{42}$$

and $\varepsilon_t^* = L\varepsilon_t$ will now have a diagonal covariance matrix. In the context of an MCMC algorithm involving $p\left(h^T|y^T, L\right)$ and $p\left(L|y^T, h^T\right)$ (where h^T stacks all $h_t = (h_{1t}', .., h_{Mt}')'$ into a $MT \times 1$ vector) we can exploit this result to run the Kim, Shephard and Chib (1998) algorithm one equation at a time. The is, conditional on an MCMC draw of L, we can transform the model as in (42) and use results for the univariate stochastic volatility model one transformed equation at a time.

Finally, to complete the MCMC algorithm for the Cogley-Sargent model we need to take draws from $p\left(L|y^T, h^T\right)$. But this is straightforward since (42) shows that this model can be written as a series of M regression equations with Normal errors which are independent of one another. Hence, standard results for the Normal linear regression model can be used to draw from $p\left(L|y^T, h^T\right)$. The appendix to Cogley and Sargent (2005) provides precise formulae for the MCMC algorithm (although their paper uses a less efficient algorithm for drawing from $p\left(h^T|y^T, L\right)$ than the algorithm in Kim, Shephard and Chib, 1998, discussed in this section).

It is worth stressing that the Cogley-Sargent model allows the covariance between the errors to change over time, but in a tightly restricted fashion related to the way the error variances are changing. This can be seen most clearly in the $M = 2$ case where ε_{1t} and ε_{2t} are the errors in the two equations. In this, case (41) implies $cov\left(\varepsilon_{1t}, \varepsilon_{2t}\right) = d_{1t} L_{21}$ which varies proportionally with the error variance of the first equation. In impulse response analysis, it can be shown that this restriction implies that a shock to the i^{th} variable has an effect on the j^{th} variable which is constant over time. In some macroeconomic applications, such a specification might be too restrictive.

Another popular approach (see, e.g., Primiceri, 2005) extends (41) to:

$$\Sigma_t = L_t^{-1} D_t L_t^{-1\prime} \tag{43}$$

where L_t is defined in the same way as L (i.e. as being a lower-triangular matrix with ones on the diagonal), but is now time varying. This specification does not restrict the covariances and variances in Σ_t in any way. The MCMC algorithm for posterior simulation from the Primiceri model is the largely same as for the model with constant L (with the trivial change that the transformation in (42) becomes $L_t y_t = L_t \varepsilon_t$). The main change in the algorithm arises in the way L_t is drawn.

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

$$l_{t+1} = l_t + \zeta_t, \tag{44}$$

where ζ_t is i.i.d. $N(0, D_\zeta)$ and independent of the other errors in the model and D_ζ is a diagonal matrix.

We have seen how the measurement equation in this model can be written as:

$$L_t y_t = \varepsilon_t^*,$$

and it can be shown that $\varepsilon_t^* \sim N(0, D_t)$. We can use the structure of L_t to isolate y_t on the left hand side and write:

$$y_t = C_t l_t + \varepsilon_t^*. \tag{45}$$

Primiceri (2005), page 845 gives a general definition of C_t. For $M = 3$,

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

where y_{it} is the i^{th} element of y_t. But (44) and (45) is now in form of a Normal linear state space model of the sort we began this section with. Accordingly, in the context of an MCMC algorithm, we can draw L_t (conditional on h^T and all the other model parameters) using an algorithm such as that of Carter and Kohn (1994) or Durbin and Koopman (2002).

Note that we have assumed D_ζ to be a diagonal matrix. Even with this restriction, the resulting multivariate stochastic volatility model is very flexible. However, should the researcher wish to have D_ζ being non-diagonal, it is worth noting that if it is simply assumed to be a positive definite matrix then the simplicity of the MCMC algorithm (i.e. allowing for the use of methods for Normal linear state space models), breaks down. Primiceri (2005) assumes D_ζ to have a certain block diagonal structure such that it is still possible to use algorithms for Normal linear state space models to draw L_t. It is also possible to extend this model to allow for D_η (the error covariance matrix in the state equation for h_t defined after equation 40) to be any positive definite matrix (rather than the diagonal one assumed previously). Exact formulae are provided in Primiceri (2005) for both these extensions. The empirical illustration below uses these generalizations of Primiceri (2005).

4 TVP-VARs

VARs are excellent tools for modeling the inter-relationships between macroeconomic variables. However, they maintain the rather strong assumption that parameters are constant over time. There are many reasons for thinking such an assumption may be too restrictive in many macroeconomic applications. Consider, for instance, U.S. monetary policy and the question of whether the high inflation and slow growth 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 monetary policy transmission mechanism. This story depends on having VAR coefficients different in the 1970s than subsequently. 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 (i.e. in the 1970s volatility was high, whereas later policymakers had the good fortune of the Great Moderation of the business cycle) which motivates the addition of multivariate stochastic volatility to VAR models . Yet others (e.g. Primiceri, 2005, Koop, Leon-Gonzalez and Strachan, 2009) have found that both the transmission mechanism and the variance of the exogenous shocks have changed over time.

This example is intended to motivate the basic point that an understanding of macroeconomic policy issues should be based on multivariate models where both the VAR coefficients and the error covariance matrix can potentially change over time. More broadly, there is a large literature in macroeconomics which documents structural breaks and other sorts of parameter change in many time series variables (see, among many others, Stock and Watson, 1996). A wide range of alternative specifications have been suggested, including Markov switching VARs (e.g. Paap and van Dijk, 2003, or Sims and Zha, 2006) and other regime-switching VARs (e.g. Koop and Potter, 2006). However, perhaps the most popular have been TVP-VARs. A very incomplete list of references which use TVP-VARs includes Canova (1993), Cogley and Sargent (2001, 2005), Primiceri (2005), Canova and Gambetti (2009), Canova and Ciccarelli (2009) and Koop, Leon-Gonzalez and Strachan (2009). In this monograph, we will not discuss regime-switching models, but rather focus on TVP-VARs.

4.1 Homoskedastic TVP-VARs

To discuss some basic issues with TVP-VAR modelling, we will begin with a homoskedastic version of the model (i.e. $\Sigma_t = \Sigma$). We will use the same definition of the dependent variables and explanatory variables as in (16) from Section 2. Remember that y_t is an $M \times 1$ vector containing data on M dependent variables and Z_t is an $M \times k$ matrix. In Section 2, we saw how Z_t could be set up to either define an unrestricted or a restricted VAR. Z_t can also contain

exogenous explanatory variables.[16] The basic TVP-VAR can be written as:

$$y_t = Z_t\beta_t + \varepsilon_t,$$

and

$$\beta_{t+1} = \beta_t + u_t, \tag{46}$$

where ε_t is i.i.d. $N(0, \Sigma)$ and u_t is i.i.d. $N(0, Q)$. ε_t and u_s are independent of one another for all s and t.

This model is similar to that used in Cogley and Sargent (2001). Bayesian inference in this model can be dealt with quite simply since it is a Normal linear state space model of the sort discussed in Section 3 of this monograph. Thus, the MCMC methods described in Section 3 can be used for Bayesian inference in the homoskedastic TVP-VAR. In many cases, this is all the researcher needs to know about TVP-VARs and Bayesian TVP-VARs work very well in practice. However, in some cases, this basic TVP-VAR can lead to poor results in practice. In the remainder of this section, we discuss how these poor results can arise and various extensions of the basic TVP-VAR which can help avoid them.

The poor results just referred to typically arise because the TVP-VAR has so many parameters to estimate. In Section 1 we saw how, even with the VAR, worries about the proliferation of parameters led to the use of priors such as the Minnesota prior or the SSVS prior. With so many parameters and relatively short macroeconomic time series, it can be hard to obtain precise estimates of coefficients. Thus, features of interest such as impulse responses can have very dispersed posterior distributions leading to wide credible intervals. Furthermore, the risk of over-fitting can be serious in some applications. In practice, it has been found that priors which exhibit shrinkage of various sorts can help mitigate these problems.

With the TVP-VAR, the proliferation of parameters problems is even more severe since it has T times as many parameters to estimate. In Section 3, we saw how the state equation in a state space model can be interpreted as a hierarchical prior (see equation 30). And, in many applications, this prior provides enough shrinkage to yield reasonable results. Although it is worth noting that it is often a good idea to use a fairly tight prior for Q. For instance, if (33) is used as a prior, then a careful choice of $\underline{\nu}_Q$ and $\underline{Q}^{-1}$ can be important in producing sensible results.[17] However, in some applications, it is desirable to introduce more prior information and we will describe several ways of doing so.

[16] The TVP-VAR where some of the coefficients are constant over time can be dealt with by adding W_t as in (28) and (29).

[17] Attempts to use flat "noninformative" priors on Q can go wrong since such flat priors actually can be quite informative, attaching large amounts of prior probability to large values of Q. Large values of Q are associated with a high degree of variation on the VAR coefficients (i.e. much prior weight is attached to regions of the parameter space where the opposite of shrinkage occurs).

4.1.1 Empirical Illustration of Bayesian Homoskedastic TVP-VAR Methods

To illustrate Bayesian inference in the homoskedastic TVP-VAR, we use the same U.S. data set as before (see Section 2.3), which contains three variables: inflation, unemployment and interest rate and runs from 1953Q1 to 2006Q3. We set lag length to 2.

We use a training sample prior of the type used by Primiceri (2005) where prior hyperparameters are set to OLS quantities calculating using a training sample of size τ (we use $\tau = 40$). Thus, data through 1962Q4 is used to choose prior hyperparameter values and then estimation uses data beginning in 1963Q1. To be precise, the training sample prior uses β_{OLS} which is the OLS estimate of the VAR coefficients in a constant-coefficient VAR and $V(\beta_{OLS})$ which is its covariance matrix.

In this model, we need a prior for the initial state, β_0, the measurement equation error covariance matrix Σ and the state equation error covariance matrix, Q. The first of these is:

$$\beta_0 \sim N\left(\beta_{OLS}, 4 \cdot V\left(\beta_{OLS}\right)\right),$$

whereas the latter two are based on (32) with $\underline{\nu} = M + 1, \underline{S} = I$ and (33) with $\underline{\nu}_Q = \tau, \underline{Q} = 0.0001 \cdot \tau \cdot V(\beta_{OLS})$.

Since the VAR regression coefficients are time-varying, there will be a different set of them in every time period. This typically will lead to far too many parameters to present in a table or graph. Here we will focus on impulse responses (defined as in Section 2.3). But even for these we have a different impulse response function at each point in time.[18] Accordingly, Figure 4 plots impulse responses for three representative times: 1975Q1, 1981Q3 and 1996Q1. For the sake of brevity, we only present the impulse responses of each variable to a monetary policy shock (i.e. a shock in the interest rate equation).

Using the prior specified above, we see that there are very small differences in these responses in three different representative periods (1975Q1, 1981Q3 and 1996Q1). In this figure the posterior median is the solid line and the dotted lines are the 10^{th} and 90^{th} percentiles. These impulse responses are, in some cases, somewhat different from the ones found in Section 2.3 for the VARs without parameter variation, which indicates the potential importance of allowing for time variation in parameters. Note also that the impulse responses in the three years differ only slightly from one another. This is consistent with a common finding by researchers using similar data sets: that much of the time variation in parameters occurs in the error covariance matrix rather than the VAR coefficients. In this model we are assuming homoskedasticity, but we will relax this assumption shortly.

[18] It is common practice to calculate the impulse responses at time time using β_t and simply ignoring the fact that it will change over time and we do so below. For more general treatments of Bayesian impulse response analysis see Koop (1996).

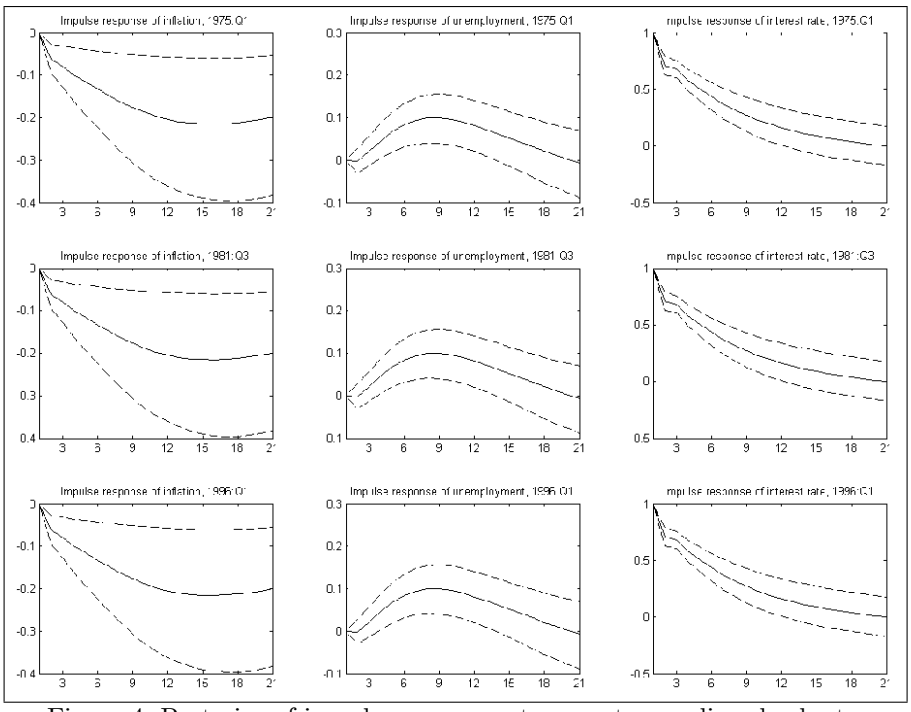

Figure 4: Posterior of impulse responses to monetary policy shock at different times

4.1.2 Combining other Priors with the TVP Prior

The Minnesota and SSVS Priors In Section 1, we described several priors that are commonly used with VARs and, when moving to TVP-VARs, an obvious thing to do is to try and combine them with the hierarchical prior defined by the state equation. This can be done in several ways. One approach, used in papers such as Ballabriga, Sebastian and Valles (1999), Canova and Ciccarelli (2004), and Canova (2007), involves combining the prior of the TVP-VAR with the Minnesota prior. This can be done by replacing (46) by

$$\beta_{t+1} = A_0\beta_t + (I - A_0)\,\overline{\beta}_0 + u_t, \tag{47}$$

where A_0 is a $k \times k$ matrix and $\overline{\beta}_0$ a $k \times 1$ vector. The matrices A_0, $\overline{\beta}_0$ and Q can either be treated as unknown parameters or specific values of them can be chosen to reflect the Minnesota prior. For instance, Canova (2007, page 399) sets $\overline{\beta}_0$ and Q to have forms based on the Minnesota prior and sets $A_0 = cI$ where c is a scalar. The reader is referred to Canova (2007) for precise details, but to provide a flavor of his recommendations, note that if $c = 1$, then the traditional TVP-VAR prior implication that $E\left(\beta_{t+1}\right) = E\left(\beta_t\right)$ is obtained, but if $c = 0$ then we have $E\left(\beta_{t+1}\right) = \overline{\beta}_0$. Canova (2007) recommends setting

the elements of $\overline{\beta}_0$ corresponding to own lags to one and all other elements to zero, thus leading to the same prior mean as in the Minnesota prior. Canova (2007) sets Q to values inspired by the prior covariance matrix of the Minnesota prior (similar to equation 8). The scalar c can either be treated as an unknown parameter or a value can be selected for it (e.g. based on a training sample).

It is also possible to treat A_0 as a matrix of unknown parameters. To carry out Bayesian inference in such a model is usually straightforward since it requires only the addition of another block in an MCMC algorithm. In Section 3.2 we saw how an MCMC algorithm involving $p\left(\Sigma^{-1}|y^T,\beta^T\right)$, $p\left(Q^{-1}|y^T,\beta^T\right)$ and $p\left(\beta^T|y^T,\Sigma,Q\right)$ could be used to carry out posterior simulation in the Normal linear state space model. If A_0 involves unknown parameters, then an MCMC algorithm involving $p\left(\Sigma^{-1}|y^T,\beta^T,A_0\right)$, $p\left(Q^{-1}|y^T,\beta^T,A_0\right)$ and $p\left(\beta^T|y^T,\Sigma,Q,A_0\right)$ proceeds in exactly the same manner. Using the notation of Section 3.2, we can simply set $T_t = A_0$ and the methods of that section can be directly used. To complete an MCMC algorithm requires draws from $p\left(A_0|y^T,\Sigma,Q,\beta^T\right)$. This will depend on the specification used for A_0, but typically this posterior conditional distribution is easy to derive using results for the VAR. That is, $p\left(A_0|y^T,\Sigma,Q,\beta^T\right)$ is a distribution which conditions on β^T, and (47) can be written in "VAR" form

$$\beta_{t+1} - \overline{\beta}_0 = A_0\left(\beta_t - \overline{\beta}_0\right) + u_t,$$

with "dependent variables" $\beta_{t+1} - \overline{\beta}_0$ and "lagged dependent variables" $\beta_t - \overline{\beta}_0$. In the context of the MCMC algorithm, these "dependent variables" and "lagged dependent variables" would be replaced with the drawn values.

Any prior can be used for $a_0 = vec(A_0)$ including any of the VAR priors described in Section 2. We will not provide exact formulae here, but note that they will be exactly as in Section 2 but with y_t replaced by $\beta_{t+1} - \overline{\beta}_0$ and x_t (or Z_t) replaced by $\beta_t - \overline{\beta}_0$.

One prior for a_0 of empirical interest is the SSVS prior of Section 2.2.3. It is interesting to consider what happens if we use the prior given in (22) and (23) for a_0 and set $\overline{\beta}_0 = 0$.[19] This implies that a_{0j} (the j^{th} element of a_0) has a prior of the form:

$$a_{0j}|\gamma_j \sim \left(1-\gamma_j\right) N\left(0,\kappa_{0j}^2\right) + \gamma_j N\left(0,\kappa_{1j}^2\right),$$

where γ_j is a dummy variable and κ_{0j}^2 is very small (so that a_{0j} is constrained to be virtually zero), but κ_{1j}^2 is large (so that a_{0j} is relatively unconstrained). The implication of combining the SSVS prior with the TVP prior is that we have a model which (with probability γ_j) says that a_{0j} is evolving according

[19] If the variables in the VAR are in levels, then the researcher may wish to set the elements of $\overline{\beta}_0$ corresponding to own first lags to one, to reflect the common Minnesota prior belief in random walk behavior.

to a random walk in the usual TVP fashion, but with probability $\left(1-\gamma_j\right)$ is set to zero. Such a model is a useful one since it allows for change in VAR coefficients over time (which is potentially of great empirical importance), but also helps avoid over-parameterization problems by allowing for some lagged dependent variables to be deleted from the VAR. Another interesting approach with a similar methodology is given in Groen, Paap and Ravazzolo (2008).

Adding Another Layer to the Prior Hierarchy Another way to combine the TVP model with prior information from another source is by adding another state equation to the TVP-VAR (i.e. another layer in the prior hierarchy). The framework that follows is taken from Chib and Greenberg (1995) and has been used in macroeconomics by, among others, Ciccarelli and Rebucci (2002).

This involves writing the TVP-VAR as:

$$\begin{aligned} y_t &= Z_t\beta_t + \varepsilon_t \\ \beta_{t+1} &= A_0\theta_{t+1} + u_t, \\ \theta_{t+1} &= \theta_t + \eta_t. \end{aligned} \tag{48}$$

where all assumptions are as for the standard TVP-VAR, but we add the assumption that η_t is i.i.d. $N\left(0,R\right)$ and η_t is independent of the other errors in the model.

Note first that there is a sense in which this specification retains random walk evolution of the VAR coefficients since it can be written as:

$$\begin{aligned} y_t &= Z_t\beta_t + \varepsilon_t \\ \beta_{t+1} &= \beta_t + v_t, \end{aligned}$$

where $v_t = A_0\eta_t + u_t - u_{t-1}$. In this sense, it is a TVP-VAR with random walk state equation but, unlike (46), the state equation errors have a particular MA(1) structure.

Another way of interpreting this model is by noting that it expresses the conditional prior belief that

$$E\left(\beta_{t+1}|\beta_t\right) = A_0\beta_t$$

and, thus, is a combination of the random walk prior belief of the conventional TVP model with the prior beliefs contained in A_0. A_0 is typically treated as known.

Note that it is possible for θ_t to be of lower dimension than β_t and this can be a useful way of making the model more parsimonious. For instance, Ciccarelli and Rebucci (2002) is a panel VAR application involving G countries and, for each country, k_G explanatory variables exist with time-varying coefficients. They specify

$$A_0 = \iota_G \otimes I_{k_G}$$

which implies that there is a time-varying component in each coefficient which is common to all countries (rather than having G different time-varying components in each country). Thus, θ_t is $k_G \times 1$ whereas β_t $k_G G \times 1$.

Posterior computation in this model is described in Chib and Greenberg (1995). Alternatively, the posterior simulation methods for state space models described in Section 3 can be used, but with a more general specification for the state equation than that given there. Just as in the preceding section, if A_0 is to be treated as a matrix of unknown parameters any prior can be used and another block added to the MCMC algorithm. The form of this block will typically be simple since, conditional on the states, we have $\beta_{t+1} = A_0\theta_{t+1} + u_t$ which has the same structure as a multivariate regression model.

As one empirically-useful example, suppose we use the SSVS prior of Section 2.2.3 for A_0. Then we obtain a model where some VAR coefficients evolve according to random walks in the standard TVP fashion while others are (approximately) omitted from the model.

4.1.3 Imposing Inequality Restrictions on the VAR Coefficients

Empirical macroeconomists typically work with multivariate time series models that they believe to be non-explosive. Thus, in TVP-VAR models it can be desirable to impose stability on the TVP-VAR at each point in time. This has lead papers such as Cogley and Sargent (2001, 2005) to restrict β_t to satisfy the usual stability conditions for VARs for $t = 1, .., T$. This involves imposing the inequality restriction that the roots of the VAR polynomial defined by β_t lie outside the unit circle. Indeed, in the absence of such a stability restriction (or a very tight prior), Bayesian TVP-VARs will place a large amount of a priori weight on explosive values for β_t (e.g. the Minnesota prior is centered over a random walk which means it allocates prior weight to the explosive region of the parameter space). This can cause problems for empirical work. For instance, even a small amount of posterior probability in explosive regions for β_t can lead to impulse responses or forecasts which have counter-intuitively large posterior means or standard deviations. Given that TVP-VARs have many parameters to estimate and the researcher often has relatively small data sets, the posterior standard deviation of β_t can often be large. Thus, even if β_t truly is stable and its posterior mean indicates stability, it is not unusual for large posterior variances to imply that appreciable posterior probability is allocated to the explosive region.

The preceding paragraph motivates one case where the researcher might wish to impose inequality restrictions on β_t in order to surmount potential over-parameterization problems which might arise in the TVP-VAR. Other inequality restrictions are also possible. In theory, imposing inequality restrictions is a good way of reducing over-parameterization problems. In practice, there is one problem with this strategy. This problem is that standard state space methods for the Normal linear state space models (see Section 3) cannot be used without some modification. Remember that MCMC methods for this model involved taking draws from $p\left(\beta^T | y^T, \Sigma, Q\right)$ and that there are many efficient algorithms

for doing so (e.g. Carter and Kohn, 1994, Fruhwirth-Schnatter, 1994, DeJong and Shephard, 1995 and Durbin and Koopman, 2002). However, all of these algorithms are derived using properties of the multivariate Normal distribution which do not carry over to the multivariate Normal distribution subject to inequality constraints.

Two methods have been proposed to impose stability restrictions (or other inequality restrictions) on TVP-VARs. These are discussed and compared in Koop and Potter (2009). The first of these involves using a standard algorithm such as that of Carter and Kohn (1994) for drawing β^T in the unrestricted VAR. If *any* drawn β_t violates the inequality restriction then the *entire* vector β^T is rejected.[20] If *every* element of β^T satisfies the inequality restriction then β^T is accepted with a certain probability (the formula for this probability is given in Koop and Potter, 2009). A potential problem with this algorithm is that it is possible for it to get stuck, rejecting virtually every β^T. In theory, if enough draws are taken this MCMC algorithm can be highly accurate, in practice "enough draws" can be so many that the algorithm simply cannot produce accurate results in a feasible amount of computer time.

In the case where no inequality restrictions are imposed, the advantage of algorithms such as that of Carter and Kohn (1994) is that they are multi-move algorithms. This means that they provide a draw for the entire vector β^T from $p\left(\beta^T|y^T,\Sigma,Q\right)$ directly. The logic of MCMC suggests that it would also be valid to draw β_t for $t=1,..,T$ one at a time from $p\left(\beta_t|y^T,\Sigma,Q,\beta_{-t}\right)$ where $\beta_{-t}=\left(\beta_1',..,\beta_{t-1},\beta_{t+1},..,\beta_T'\right)'$. It is indeed the case that this is also valid thing to do. However, it is rarely done in practice since such single-move algorithms will be slow to mix. That is, they will tend to produce a highly correlated series of draws which means that, relative to multi-move algorithms, more draws must be taken to achieve a desired level of accuracy. The second algorithm proposed for the TVP-VAR subject to inequality restrictions is a single-move algorithm. This algorithm does have the disadvantage just noted – that it is slow mixing. But it is possible that this disadvantage is outweighed by the advantage that the single-move algorithm does not run into the problem noted above for the multi-move algorithm (i.e. that the multi-move algorithm can get stuck and reject every draw).

Koop and Potter (2009) provide full details of both these algorithms and weigh their advantages and disadvantages in a macroeconomic application.

4.1.4 Dynamic Mixture Models

Another way of tightening the parameterization of the TVP-VAR is through the dynamic mixture model approach of Gerlach, Carter and Kohn (2000). This has recently been used in models of relevance for empirical macroeconomics in Giordani, Kohn and van Dijk (2007) and Giordani and Kohn (2008) and applied to TVP-VARs in Koop, Leon-Gonzalez and Strachan (2009).

[20] It can be shown that the strategy of rejecting only individual β_t which violate the inequality restriction is not a valid one.

To explain the usefulness of dynamic mixture modelling for TVP-VAR analysis, return to the general form for the Normal linear state space model given in (28) and (29) and remember that this model depended on so-called system matrices, Z_t, Q_t, T_t, W_t and Σ_t. Dynamic mixture models allow for any or all of these system matrices to depend on an $s \times 1$ vector $\widetilde{K}_t$. Gerlach, Carter and Kohn (2000) discuss how this specification results in a mixtures of Normals representation for y_t and, hence, the terminology dynamic mixture (or mixture innovation) model arises. The contribution of Gerlach, Carter and Kohn (2000) is to develop an efficient algorithm for posterior simulation for this class of models. The efficiency gains occur since the states are integrated out and $\widetilde{K} = \left(\widetilde{K}_1, .., \widetilde{K}_T\right)'$ is drawn unconditionally (i.e. not conditional on the states). A simple alternative algorithm would involve drawing from the posterior for $\widetilde{K}$ conditional on β^T (and the posterior for β^T conditional on $\widetilde{K}$). Such a strategy can be shown to produce a chain of draws which is very slow to mix. The Gerlach, Carter and Kohn (2000) algorithm requires only that $\widetilde{K}_t$ be Markov (i.e. $p\left(\widetilde{K}_t|\widetilde{K}_{t-1}, .., \widetilde{K}_1\right) = p\left(\widetilde{K}_t|\widetilde{K}_{t-1}\right)$) and is particularly simple if $\widetilde{K}_t$ is a discrete random variable. We will not provide details of their algorithm here, but refer the reader to Gerlach, Carter and Kohn (2000) or Giordani and Kohn (2008). This algorithm is available on the Matlab website associated with this monograph.

The dynamic mixture framework can be used in many ways in empirical macroeconomics, here we illustrate one useful way. Consider the following TVP-VAR:

$$y_t = Z_t\beta_t + \varepsilon_t,$$

and

$$\beta_{t+1} = \beta_t + u_t,$$

where ε_t is i.i.d. $N(0, \Sigma)$ and u_t is i.i.d. $N\left(0, \widetilde{K}_t Q\right)$. This model is exactly the same as the TVP-VAR of Section 4.1, except for the error covariance matrix in the state equation. Let $\widetilde{K}_t \in \{0, 1\}$ and assume a hierarchical prior for it of the following form:

$$\begin{aligned} p\left(\widetilde{K}_t = 1\right) &= q. \\ p\left(\widetilde{K}_t = 0\right) &= 1 - q \end{aligned}$$

where q is an unknown parameter.[21]

This is a simple example of a dynamic mixture model. It has the property that:

[21] This specification for $\widetilde{K}_t$ is Markov in the degenerate sense that an independent process is a special case of Markov process.

$$\begin{array}{l} \beta_{t+1} = \beta_t + u_t \text{ if } \widetilde{K}_t = 1 \\ \beta_{t+1} = \beta_t \text{ if } \widetilde{K}_t = 0 \end{array} .$$

In words, the VAR coefficients can change at time t (if $\widetilde{K}_t = 1$), but can also remain constant (if $\widetilde{K}_t = 0$) and q is the probability that the coefficients change. $\widetilde{K}_t$ and q are estimated in a data-based fashion. Thus, this model can have the full flexibility of the TVP-VAR if the data warrant it (in the sense that it can select $\widetilde{K}_t = 1$ for $t = 1, .., T$). But it can also select a much more parsimonious representation. In the extreme, if $\widetilde{K}_t = 0$ for $t = 1, .., T$ then we have the VAR without time-varying parameters.

This simple example allows either for all the VAR coefficients to change at time t (if $\widetilde{K}_t = 1$) or none (if $\widetilde{K}_t = 0$). More sophisticated models can allow for some parameters to change but not others. For instance, $\widetilde{K}_t$ could be a vector of M elements, each being applied to one of the equations in the VAR. Such a model would allow the VAR coefficients in some equations to change, but remain constant in other equations. In models with multivariate stochastic volatility, $\widetilde{K}_t$ could contain elements which control the variation in the measurement error covariance matrix, Σ_t. This avenue is pursued in Koop, Leon-Gonzalez and Strachan (2009). Many other possibilities exist and the advantage of the dynamic mixture framework is that there exists a well-developed, well-understood set of MCMC algorithms that make Bayesian inference straightforward.

4.2 TVP-VARs with Stochastic Volatility

Thus far, we have focussed on the homoskedastic TVP-VAR, assuming the error covariance matrix, Σ to be constant. However, we have argued above (see Section 3.3) that volatility issues are often very important in empirical macroeconomics. Thus, in most cases, it is important to allow for multivariate stochastic volatility in the TVP-VAR. Section 3.3.2 discusses multivariate stochastic volatility, noting that there are many possible specifications for Σ_t. Particular approaches used in Cogley and Sargent (2005) and Primiceri (2005) are described in detail. All that we have to note here is that either of these approaches (or any other alternative) can be added to the homoskedastic TVP-VAR.

With regards to Bayesian inference using MCMC methods, we need only add another block to our algorithm to draw Σ_t for $t = 1, .., T$. That is, with the homoskedastic TVP-VAR we saw how an MCMC algorithm involving $p\left(Q^{-1}|y^T, \beta^T\right)$, $p\left(\beta^T|y^T, \Sigma, Q\right)$ and $p\left(\Sigma^{-1}|y^T, \beta^T\right)$ could be used. When adding multivariate stochastic volatility, the first of these densities is unchanged. The second, $p\left(\beta^T|y^T, \Sigma, Q\right)$, becomes $p\left(\beta^T|y^T, \Sigma_1, .., \Sigma_T, Q\right)$ which can be drawn from using any of the algorithms (e.g. Carter and Kohn, 1994) for the Normal linear state space model mentioned previously. Finally, $p\left(\Sigma^{-1}|y^T, \beta^T\right)$ is replaced by $p\left(\Sigma_1^{-1}, .., \Sigma_T^{-1}|y^T, \beta^T\right)$. Draws from this posterior conditional can be taken as described in Section 3.3.2.

4.3 Empirical Illustration of Bayesian Inference in TVP-VARs with Stochastic Volatility

We continue our empirical illustration of Section 4.1.1 which involved a homoskedastic TVP-VAR. All details of the empirical are as in Section 4.1.1 except we additionally allow for multivariate stochastic volatility as in Primiceri (2005). The prior for the parameters relating to the multivariate stochastic volatility are specified as in Primiceri (2005).

For the sake of brevity, we do not present impulse responses (these are similar to those presented in Primiceri, 2005). Instead we present information relating to the multivariate stochastic volatility. Figure 6 presents the time-varying standard deviations of the errors in the three equations of our TVP-VAR (i.e. the posterior means of the square roots of the diagonal element of Σ_t). Figure 5 shows that there is substantial time variation in the error variances in all equations. In particular, it can be seen that 1970s was a very volatile period for the US economy, while the monetarist experiment of the early 1980s is also with instability. However, after the early 1980s volatility is greatly reduced in all equations. This latter period has come to be known as the Great Moderation of the business cycle.

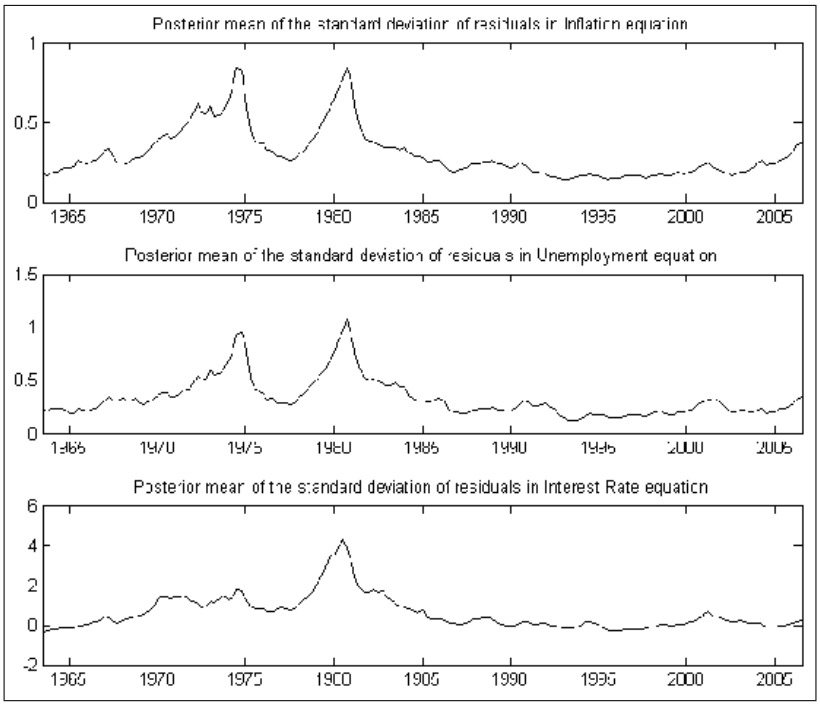

Figure 5: Time-varying volatilities of errors in the three equations of the TVP-VAR

5 Factor Methods

The VAR and TVP-VAR methods we have discussed so far are typically used when the number of variables in the model is relatively small (e.g. three or four and rarely more than ten).[22] However, in most modern economies, the macroeconomist will potentially have dozens or hundreds of time series variables to work with. Especially when forecasting, the researcher wants to include as much information as possible, and it can be desirable to work with as many variables as possible. This can lead to models with a large number of variables and a proliferation of parameters. Accordingly, researchers have sought ways of extracting the information in data sets with many variables while, at the same time, keeping the model parsimonious. Beginning with Geweke (1977) factor models have been the most common way of achieving this goal (see Stock and Watson 2006 for a recent survey). Applications such as Forni and Reichlin (1998), Stock and Watson (1999, 2002), Bernanke and Boivin (2003), have popularized factor methods among macroeconomists and Geweke and Zhou (1996), Otrok and Whiteman (1998) and Kose, Otrok and Whiteman (2003), among many others, stimulated the interest of Bayesians. Papers such as Bernanke, Boivin and Eliasz (2005) and Stock and Watson (2005) have combined factor methods with VAR methods. More recently, papers such as Del Negro and Otrok (2008) and Korobilis (2009a) provide further TVP extensions to these models. In this section, we will describe dynamic factor models and their extensions to factor augmented VARs (FAVARs) and TVP-FAVARs. As we shall see, these models can be interpreted as state space models and Bayesian MCMC algorithms that use our previously-discussed algorithms for state space models can be used.

5.1 Introduction

We will retain our notation where y_t is an $M \times 1$ vector of time series variables, but now M will be very large and let y_{it} denote a particular variable. A simple static factor model is (see Lopes and West, 2004):

$$y_t = \lambda_0 + \lambda f_t + \varepsilon_t, \tag{49}$$

The key aspect of the factor model is the introduction of f_t which is a $q \times 1$ vector of unobserved latent factors (where $q << M$) which contains information extracted from all the M variables. The factors are common to every dependent variable (i.e. the same f_t occurs in every equation for y_{it} for $i = 1, ..., M$), but they may have different coefficients (λ which is a $M \times q$ matrix of so-called factor loadings). Also, the equation for every dependent variable has its own intercept (i.e. λ_0 is $M \times 1$ vector of parameters). ε_t is i.i.d. $N(0, \Sigma)$. Different factor models arise from the assumptions about the factors. For example, the simplest case would be to assume that the factors come from a standard Normal density, $f_t \sim N(0, I)$. This implies that the covariance matrix of the observed data can

[22] An important exception is Banbura, Giannone and Reichlin (2008) which uses Bayesian VARs (with time-invariant coefficients) with up to 130 variables

be written as:

$$var(y) = \lambda\lambda' + \Sigma.$$

Alternatively, if we assume that the factors have a (not necessarily diagonal) covariance matrix Σ^f, the decomposition becomes

$$var(y) = \lambda\Sigma^f\lambda' + \Sigma.$$

Even in this simple static framework, many extensions are possible. For example, Pitt and Shephard (1999) assume a factor stochastic volatility specification (i.e. a diagonal factor covariance matrix Σ_t^f which varies over time, with diagonal elements following a geometric random walk). West (2003) uses SSVS on the parameters (λ_0, λ).

We write out these covariance matrices to illustrate the identification issues which arise in factor models. In general, $var(y)$ will have $\frac{M(M+1)}{2}$ elements which can be estimated. However, without further restrictions, Σ and λ (or Σ, Σ_f and λ) will have many more elements than this. It is common to restrict Σ to be a diagonal matrix. This restriction implies that all the commonalities across variables occur through the factors and that the individual elements of ε_t are purely shocks which are idiosyncratic to each variable. But additional identifying restrictions are typically required. Lopes and West (2004) and Geweke and Zhou (1996) give a clear explanation of why identification is not achieved in the simple factor model. Below we will discuss more flexible factor models, but we stress that this added flexibility requires the imposition of more identification restrictions. There is not a single, universally agreed-upon method for achieving identification in these models. Below, we make particular choices that have been used in the literature, but note that others are possible.

5.2 The Dynamic Factor Model

In macroeconomics, it is common to extend the static factor model to allow for the dynamic properties which characterize macroeconomic variables. This leads us to dynamic factor models (DFMs). A popular DFM is:

$$\begin{array}{l} y_{it} = \lambda_{0i} + \lambda_i f_t + \varepsilon_{it} \\ f_t = \Phi_1 f_{t-1} + .. + \Phi_p f_{t-p} + \varepsilon_t^f \\ \varepsilon_{it} = \phi_{i1}\varepsilon_{it-1} + .. + \phi_{ip_i}\varepsilon_{it-p_i} + u_{it} \end{array} , \tag{50}$$

where f_t is defined as in (49), λ_i which is a $1 \times q$ vector of factor loadings. Also, the equation for every dependent variable has its own intercept, λ_{0i}. The error in each equation, ε_{it}, may be autocorrelated, as specified in the third equation in (50) which assumes u_{it} to be i.i.d. $N\left(0, \sigma_i^2\right)$. The vector of factors is assumed to follow a VAR process with ε_t^f being i.i.d. $N\left(0, \Sigma^f\right)$. The errors, u_{it} are independent over i, t and of ε_t^f.

Many slight modifications of the DFM given in (50) have been used in the literature. But this specification is a popular one so we will discuss factor models using this framework. Note that it incorporates many identifying assumptions

and other assumptions to ensure parsimony. In particular, the assumption that $\Sigma^f = I$ is a common identifying assumption.[23] If Σ^f is left unrestricted, then it is impossible to separately identify λ_i and Σ^f. Similarly, the identifying assumption that u_{it} is uncorrelated with u_{jt} (for $i \neq j$) and ε_t^f is a standard one which assures that the co-movements in the different variables in y_t arises from the factors. Even with these assumptions, there is a lack of identification since the first equation of (50) can be replaced by:

$$y_{it} = \lambda_{0i} + \lambda_i C' C f_t + \varepsilon_{it},$$

where C is any orthonormal matrix.[24] Thus, (50) is observationally equivalent to a model with factors Cf_t and factor loadings $\lambda_i C'$. One way of surmounting this issue is to impose restrictions on λ_i (as is done below in our empirical example). A further discussion of various DFMs and identification is provided, e.g., in Stock and Watson (2005). See also Sentana and Fiorentini (2001) for a deeper discussion of identification issues.

To simplify the following discussion, we will assume the errors in the measurement equation are not autocorrelated (i.e. ε_{it} are i.i.d. $N\left(0, \sigma_i^2\right)$ and, thus,$\phi_{i1} = .. = \phi_{ip_i} = 0$). We do this not because the extension to autocorrelated errors is empirically unimportant (it may be important), but because it involves a straightforward addition of other blocks to the MCMC algorithm of a standard form. That is, adding AR (or ARMA) errors to a regression model such as the first equation of (50) involves standard methods (see, e.g., Chib and Greenberg, 1994). To be precise, the Chib and Greenberg (1994) algorithm will produce draws of $\phi_{i1}, .., \phi_{ip}$ and these can be plugged into the usual quasi-differencing operator for AR models and this operator can be applied to the first equation in (50). The methods described below can then be used to draw all the other parameters of this model, except that y_{it} and f_t will be replaced by quasi-differenced versions.

5.2.1 Replacing Factors by Estimates: Principal Components

Before discussing a full Bayesian analysis of the DFM which (correctly) treats f_t as a vector of unobserved latent variables, it is worth noting a simple approximation that may be convenient in practice. This approximation is used, e.g., in Koop and Potter (2004). It involves noting that the DFM has approximately the same structure as the regression model:

$$y_{it} = \lambda_{0i} + \widetilde{\lambda}_{0i} f_t + .. + \widetilde{\lambda}_{pi} f_{t-p} + \widetilde{\varepsilon}_{it}. \tag{51}$$

Thus, if f_t were known we could use Bayesian methods for the multivariate Normal regression model to estimate or forecast with the DFM.

[23] However, if (as below and in Pitt and Shephard, 1999) the researcher wishes to allow for factor stochastic volatility then this identifying assumption cannot be made and an alternative one is necessary.

[24] An orthonormal matrix has the property that $C'C = I$.

However, it is common to use principal components methods to approximate f_t.[25] Hence, approximate methods for Bayesian analysis of the DFM can be carried out by simply replacing f_t (51) by a principal components estimate and using regression methods.

5.2.2 Treating Factors as Unobserved Latent Variables

It is also not difficult to treat the factors as unobserved latent variables using Bayesian methods for state space models that were discussed in Section 3. This is particularly easy to see if we ignore the AR structure of the errors in the measurement equation and write the DFM as

$$\begin{aligned} y_{it} &= \lambda_{0i} + \lambda_i f_t + \varepsilon_{it} \\ f_t &= \Phi_1 f_{t-1} + .. + \Phi_p f_{t-p} + \varepsilon_t^f \end{aligned} \tag{52}$$

for $i = 1, .., M$ where ε_{it} is i.i.d. $N\left(0, \sigma_i^2\right)$ and ε_t^f is i.i.d. $N\left(0, \Sigma^f\right)$. In this form it can be seen clearly that the DFM is a Normal linear state space model of the form given in (28) and (29). Thus all the methods for posterior simulation introduced in Section 3 can be used to carry out Bayesian inference. In the following we provide some additional detail about the steps involved.

Note first that, conditional on the model's parameters, $\Sigma^f, \Phi_1, .., \Phi_p, \lambda_{0i}, \lambda_i, \sigma_i^2$ for $i = 1, .., M$, any of the standard algorithms for state space models such as that of Carter and Kohn (1994) can be used to draw the factors. But conditional on the factors, the measurement equations are just M Normal linear regression models. Note that the assumption that ε_{it} is independent of ε_{jt} for $i \neq j$ means that the posteriors for $\lambda_{0i}, \lambda_i, \sigma_i^2$ in the M equations are independent over i and, hence, the parameters for each equation can be drawn one at a time. Finally, conditional on the factors, the state equation becomes a VAR and the methods for Bayesian analysis in VARs of Section 2 can be used. Details on the derivations for this or related models can be found in many places, including Geweke and Zhou (1996), Kim and Nelson (1999), Lopes and West (2004) or Del Negro and Schorfheide (2009). Lopes and West (2004) discusses the choice of q, the number of factors.

5.2.3 Impulse Response Analysis in the DFM

In our discussion of impulse response analysis in VAR models, we emphasized how it is typically done based on a structural VAR:

$$C_0 y_t = c_0 + \sum_{j=1}^{p} C_j y_{t-j} + u_t$$

[25] If Y is the $T \times M$ matrix containing all the variables and W is a $M \times q$ matrix containing the eigenvectors corresponding to the q largest eigenvalues of $Y'Y$, then $F = YW$ produces an estimate of the matrix of the factors.

or $C(L) y_t = c_0 + u_t$ where u_t is i.i.d. $N(0, I)$, $C(L) = C_0 - \sum_{j=1}^{p} C_j L^p$ and L is the lag operator. In fact, impulse responses are coefficients in the vector moving averaging (VMA) representation:

$$y_t = \widetilde{c}_0 + C(L)^{-1} u_t,$$

where $\widetilde{c}_0$ is an intercept ($\widetilde{c}_0 = C(L)^{-1} c_0$). To be precise, the response of the i^{th} variable to the j^{th} structural error h periods in the future, will be the $(ij)^{th}$ element of the VMA coefficient matrix on u_{t-h}. This makes clear that what is required for impulse response analysis is a VMA representation and a method for structurally identifying shocks through a choice of C_0.

With the DFM, we can obtain the VMA representation for y_t. By substituting the state equation in (52) into the measurement equation we can obtain:

$$\begin{aligned} y_t &= \varepsilon_t + \lambda \Phi(L)^{-1} \varepsilon_t^f \\ &= B(L) \eta_t \end{aligned}$$

where $\Phi(L) = I - \Phi_1 L - .. - \Phi_p L^p$ and, for notational simplicity, we have suppressed the intercept and are still assuming ε_t to be serially uncorrelated. Adding such extensions is conceptually straightforward (e.g. implying $\varepsilon_{t-1}, .., \varepsilon_{t-p}$ would be included in the VMA representation).

In the VMA form it can be seen that standard approaches to impulse response analysis run into trouble when the DFM is used since the errors in the VMA are a combination of the measurement equation errors, ε_t, and the state equation errors, ε_t^f. For instance, in the VAR it is common to identify the structural shocks by assuming C_0 to be a lower-triangular matrix. If the interest rate were the last element of y_t this would ensure that the error in the interest rate equation had no immediate effect on the other variables thus identifying it as a monetary policy shock under control of the central bank (i.e. the shock will be proportional to the change in interest rates). With the DFM, a monetary policy shock defined by assuming C_0 to be lower triangular will not purely reflect the interest rate change, but will reflect the change in the interest rate and relevant element of ε_t^f. Thus, impulse response analysis in DFMs is problematic: it is difficult to identify an economically-sensible structural shock to measure impulse responses to. This motivates the use of FAVARs. In one sense, these are simply a different form for writing the DFM but implicitly involve a restriction which allows for economically-sensible impulse response analysis.[26]

5.3 The Factor Augmented VAR (FAVAR)

DFMs are commonly used for forecasting. However, interest in combining the theoretical insights provided by VARs with factor methods' ability to extract information in large data sets motivates the development of factor augmented

[26] This is the interpretation of the FAVAR given by Stock and Watson (2005) who begin by writing the FAVAR as simply being a DFM written in VAR form.

VARs or FAVARs. For instance, a VAR with an identifying assumption which isolates a monetary policy shock can be used to calculate impulse responses which measure the effect of monetary policy. As we have seen, such theoretical insights are hard to obtain in the DFM. However, VARs typically involve only a few variables and it is possible that this means important economic information is excluded.[27] This suggests that combining factor methods, which extract the information in hundreds of variables, with VAR methods might be productive. This is done in papers such as Bernanke, Boivin and Eliasz (2005) and Belviso and Milani (2006).

The FAVAR modifies a DFM such as (52) by adding other explanatory variables to the M measurement equations:

$$y_{it} = \lambda_{0i} + \lambda_i f_t + \gamma_i r_t + \varepsilon_{it}, \tag{53}$$

where r_t is a $k_r \times 1$ vector of observed variables. For instance, Bernanke, Boivin and Eliasz (2005) set r_t to be the Fed Funds rate (a monetary policy instrument) and, thus, $k_r = 1$. All other assumptions about the measurement equation are the same as for the DFM.

The FAVAR extends the state equation for the factors to also allow for r_t to have a VAR form. In particular, the state equation becomes:

$$\begin{pmatrix} f_t \\ r_t \end{pmatrix} = \widetilde{\Phi}_1 \begin{pmatrix} f_{t-1} \\ r_{t-1} \end{pmatrix} + .. + \widetilde{\Phi}_p \begin{pmatrix} f_{t-p} \\ r_{t-p} \end{pmatrix} + \widetilde{\varepsilon}_t^f \tag{54}$$

where all state equation assumptions are the same as for the DFM with the extension that $\widetilde{\varepsilon}_t^f$ is i.i.d. $N\left(0, \widetilde{\Sigma}^f\right)$.

We will not describe the MCMC algorithm for carrying out Bayesian inference in the FAVAR since it is very similar to that for the DFM.[28] That is, (53) and (54) is a Normal linear state space model and, thus, standard methods (e.g. from Carter and Kohn, 1994) described in Section 3 can be used to draw the latent factors (conditional on all other model parameters). Conditional on the factors, the measurement equations are simply univariate Normal linear regression models for which Bayesian inference is standard. Finally, conditional on the factors, (54) is a VAR for which Bayesian methods have been discussed in this monograph.

5.3.1 Impulse Response Analysis in the FAVAR

With the FAVAR, impulse responses of all the variables in y_t to the shocks associated with r_t can be calculated using standard methods. For instance, if r_t is the interest rate and, thus, the error in its equation is the monetary policy

[27] As an example, VARs with a small number of variables sometimes lead to counter-intuitive impulse responses such as the commonly noted price puzzle (e.g. where increases in interest rates seem to increase inflation). Such puzzles often vanish when more variables are included in the VAR suggesting that VARs with small numbers of variables may be mis-specified.

[28] The working paper version of Bernanke, Boivin and Eliasz (2005) has an appendix which provides complete details. See also the Matlab manual on the website associated with this monograph.

shock, then the response of any of the variables in y_t to the monetary policy shock can be calculated using similar methods as for the VAR. To see this note that the FAVAR model can be written as:

$$\begin{aligned}\begin{pmatrix} y_t \\ r_t \end{pmatrix} &= \begin{bmatrix} \lambda & \gamma \\ 0 & 1 \end{bmatrix} \begin{pmatrix} f_t \\ r_t \end{pmatrix} + \widetilde{\varepsilon}_t, \\ \begin{pmatrix} f_t \\ r_t \end{pmatrix} &= \widetilde{\Phi}_1 \begin{pmatrix} f_{t-1} \\ r_{t-1} \end{pmatrix} + .. + \widetilde{\Phi}_p \begin{pmatrix} f_{t-p} \\ r_{t-p} \end{pmatrix} + \widetilde{\varepsilon}_t^f\end{aligned}$$

where $\widetilde{\varepsilon}_t = (\varepsilon_t', 0)'$ and γ is an $M \times k_r$ matrix containing the γ_is. As for the DFM model, for notational simplicity, we have suppressed the intercept and assumed ε_t to be serially uncorrelated. Adding such extensions is straightforward.

If we write the second equation in VMA form as $\begin{pmatrix} f_t \\ r_t \end{pmatrix} = \widetilde{\Phi}(L)^{-1} \widetilde{\varepsilon}_t^f$ (where $\widetilde{\Phi}(L) = I - \widetilde{\Phi}_1 L - .. - \widetilde{\Phi} L^p$) and substitute into the first equation, we obtain:

$$\begin{aligned}\begin{pmatrix} y_t \\ r_t \end{pmatrix} &= \begin{bmatrix} \lambda & \gamma \\ 0 & 1 \end{bmatrix} \widetilde{\Phi}(L)^{-1} \widetilde{\varepsilon}_t^f + \widetilde{\varepsilon}_t \\ &= \widetilde{B}(L) \eta_t.\end{aligned}$$

Thus, we have a VMA form which can be used for impulse response analysis. But consider the last k_r elements of η_t which will be associated with the equations for r_t. Unlike with the DFM, these VMA errors are purely the errors associated with the VAR for r_t. This can be seen by noting that the last k_r elements of $\widetilde{\varepsilon}_t$ are zero and thus the corresponding elements of η_t will only reflect corresponding elements of $\widetilde{\varepsilon}_t^f$ which are errors in equations having r_t as dependent variables. Unlike in the DFM, they do not combine state equation errors with measurement equation errors. For instance, if r_t is an interest rate and structural identification is achieved by assuming C_0 (see equation 14) to be lower-triangular, then the structural shock to the interest rate equation is truly proportional to a change in the interest rate and the response to such a monetary policy shock has an economically-sensible interpretation.

Remember that, as with any factor model, we require identification restrictions (e.g. principal components methods implicitly involve an identification restriction that the factors are orthogonal, but other restrictions are possible). In order to do structural impulse response analysis, additional identification restrictions are required (e.g. that C_0 is lower-triangular). Note also that the restrictions such as C_0 being lower triangular are timing restrictions and must be thought about carefully. For instance, Bernanke, Boivin and Eliasz (2005) divide the elements of y_t into blocks of "slow variables" (i.e. those which are slow to respond to a monetary policy shock) and "fast variables" as part of their identification scheme.

5.4 The TVP-FAVAR

In this monograph, we began by discussing Bayesian VAR modelling, before arguing that it might be desirable to allow the VAR coefficients to vary over time. This led us to the homoskedastic TVP-VAR. Next, we argued that it is usually empirically important to allow for multivariate stochastic volatility. We can go through exactly the same steps with the FAVAR resulting in a TVP-FAVAR. Several specifications have been proposed, including Del Negro and Otrok (2008) and Korobilis (2009a). However, just as with TVP-VARs, it is worth stressing that TVP-FAVARs can be over-parameterized and that careful incorporation of prior information or the imposing of restrictions (e.g. only allowing some parameters to vary over time) can be important in obtaining sensible results. Completely unrestricted versions of them can be difficult to estimate using MCMC methods since they involve so many different state equations (i.e. one set for the factors and others to model the evolution of the parameters).

A very general specification for the TVP-FAVAR is given in Korobilis (2009a)[29] who replaces (53) and (54) by

$$y_{it} = \lambda_{0it} + \lambda_{it} f_t + \gamma_{it} r_t + \varepsilon_{it}, \tag{55}$$

and

$$\begin{pmatrix} f_t \\ r_t \end{pmatrix} = \widetilde{\Phi}_{1t} \begin{pmatrix} f_{t-1} \\ r_{t-1} \end{pmatrix} + .. + \widetilde{\Phi}_{pt} \begin{pmatrix} f_{t-p} \\ r_{t-p} \end{pmatrix} + \widetilde{\varepsilon}_t^f \tag{56}$$

and assumes each ε_{it} follows a univariate stochastic volatility process and $var\left(\widetilde{\varepsilon}_t^f\right) = \widetilde{\Sigma}_t^f$ has a multivariate stochastic volatility process of the form used in Primiceri (2005). Finally, the coefficients (for $i = 1, .., M$) $\lambda_{0it}, \lambda_{it}, \gamma_{it}, \widetilde{\Phi}_{1t}, .., \widetilde{\Phi}_{pt}$ are allowed to evolve according to random walks (i.e. state equations of the same form as 46 complete the model). All other assumptions are the same as for the FAVAR.

We will not describe the MCMC algorithm for this model other than to note that it simply adds more blocks to the MCMC algorithm for the FAVAR. These blocks are all of forms previously discussed in this monograph. For instance, the error variances in the measurement equations are drawn using the univariate stochastic volatility algorithm of Section 3.3.1. The algorithm of Section 3.3.2 can be used to draw $\widetilde{\Sigma}_t^f$. The coefficients $\lambda_{0it}, \lambda_{it}, \gamma_{it}, \widetilde{\Phi}_{1t}, .., \widetilde{\Phi}_{pt}$ are all drawn using the algorithm of Section 3.2, implemented in a very similar fashion as in the TVP-VAR. In short, as with so many models in empirical macroeconomics, Bayesian inference in the TVP-FAVAR proceeds by putting together an MCMC algorithm involving blocks from several simple and familiar algorithms.

[29] The model of Korobilis (2009a) is actually a slight extension of this since it includes a dynamic mixture aspect similar to that presented in Section 4.1.3.

5.5 Empirical Illustration of Factor Methods

To illustrate Bayesian inference in FAVARs and TVP-FAVAR models, we use a data set of 115 quarterly US macroeconomic variables spanning 1959Q1 though 2006Q3. Following common practice in this literature, we transform all variables to be stationary. For brevity, we do not list these variables here nor describe the required stationarity transformations. These details are provided in the manual on the website associated with this monograph.

The FAVAR is given in (53) and (54). It requires the choice of variables to isolate in r_t and we use the same variables as in our previous VAR and TVP-VAR empirical illustrations: inflation, unemployment and the interest rate. Consequently, our FAVAR is the same tri-variate VAR used in previous empirical illustrations, augmented with factors, f_t, which are extracted from a large set of macroeconomic and financial variables.

We use principal components methods to extract the first two factors which are used in the FAVAR ($q = 2$) and two lags in the factor equation ($p = 2$).[30] The use of principal components methods ensures identification of the model since it normalizes all factors to have mean zero and variance one. For the FAVAR we require a prior for the parameters $\Sigma^f, \Phi_1, .., \Phi_p, \lambda_{0i}, \lambda_i, \sigma_i^2$ for $i = 1, .., M$. Full details of this (relatively noninformative) prior are provided in the manual on the website associated with this monograph.

To carry out impulse response analysis we require additional identifying assumptions. With regards to the equations for r_t, we adopt the same identifying assumptions as in our previous empirical illustrations. These allow us to identify a monetary policy shock. With regards to the variables in y_t, suffice it to note here is that we adopt the same assumptions as Bernanke, Boivin and Eliasz (2005). The basic idea of their identifying scheme is described in Section 5.2.3 above.

Figures 6 and 7 plot impulse responses to the monetary policy shock for the FAVAR. Figure 6 plots impulse responses for the main variables which are included in r_t. The patterns in Figure 6 are broadly similar as those obtained in our empirical illustration using VARs (compare Figure 6 to Figure 4). However, the magnitudes of the impulse responses are somewhat different. Here we are finding more evidence that a monetary policy shock will decrease inflation. Figure 7 plots the response of a few randomly selected variables to the monetary policy shock.[31]

[30] These choices are only illustrative. In a substantive empirical exercise, the researcher would select these more carefully (e.g. using model selection methods such discussed previously in this monograph).

[31] See the manual on the website associated with this monograph for a definition of the abbreviations used in Figures 7 and 10. Briefly GDPC96 is GDP, GSAVE is savings, PRFI is private residential fixed investment, MANEMP is employment in manufacturing, AHEMAN is earnings in manufacturing, HOUST is housing starts, GS10 is a 10 year interest rate, EXJPUS is the Japanese-US exchange rate, PPIACO is a producer price index, OILPRICE is the oil price, HHSNTN is an index of consumer expectations and PMNO is the NAPM orders index. All impulse responses are to the original untransformed versions of these variables.

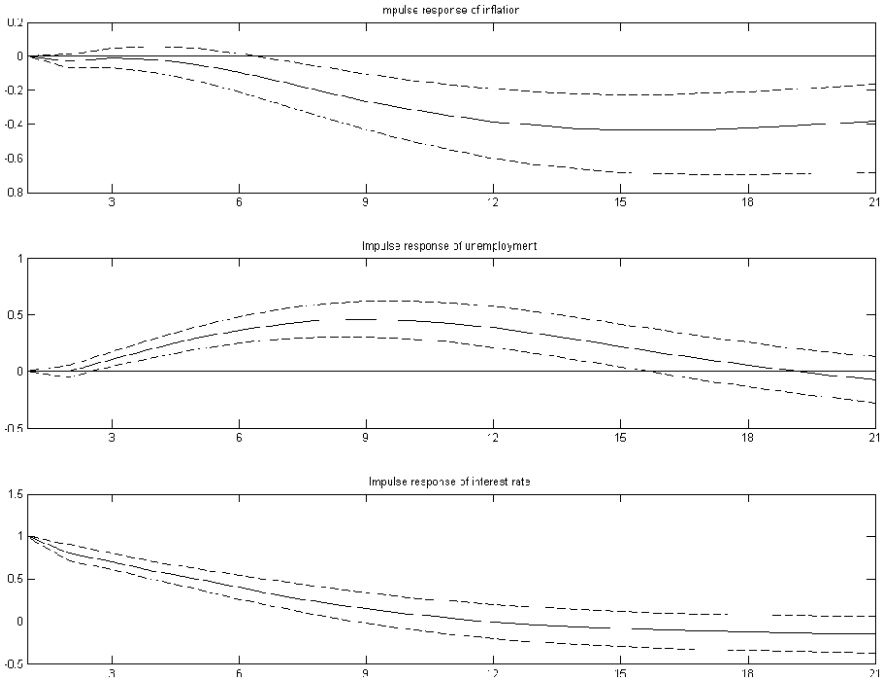

Figure 6: Posterior of impulse responses of main variables to monetary policy shock

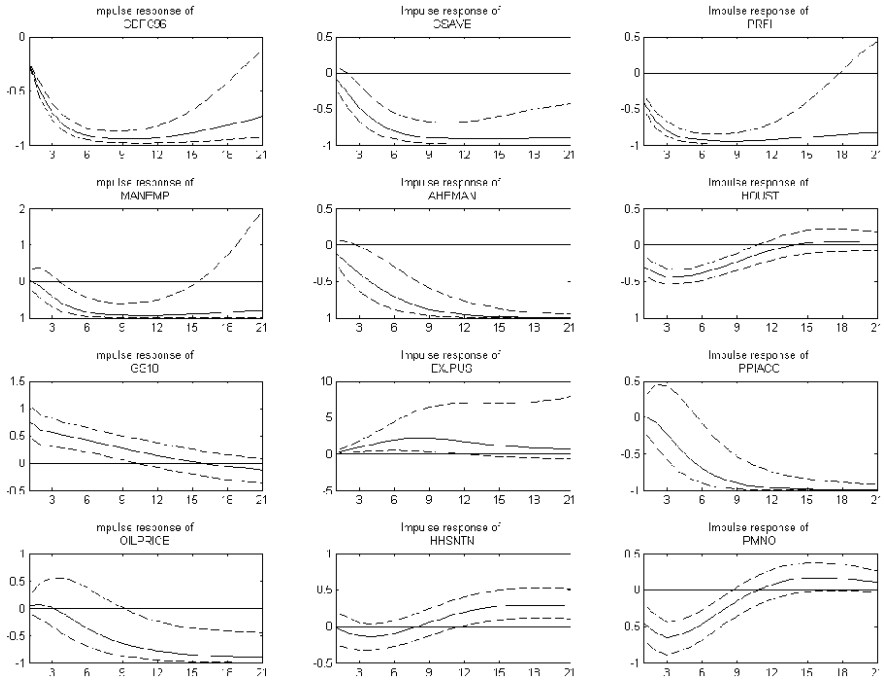

Figure 7: Posterior of impulse responses of selected variables to monetary policy shock

We now present results for a TVP-FAVAR defined in (55) and (56), but with the restriction that the parameters in the measurement equation are constant over time.[32] The (relatively noninformative) prior we use is described in the manual on the website associated with this monograph.

Figure 8 plots the posterior means of the standard deviations of the errors in the two equations where the factors are the dependent variable and the three equations where r_t are the dependent variables. It can be seen that there is substantive evidence of variation in volatilities (in particular, for equation for the first factor). The bottom three panels of Figure 8 look similar to Figure 5 and indicate the substantial increase in volatility associated with the 1970s and early 1980s preceding the Great Moderation of the business cycle. However, this pattern is somewhat muted relative to Figure 5 since the inclusion of the factors means that the standard deviation of the errors becomes smaller.

[32] We do this since it is difficult to estimate time variation in coefficients in both the measurement and state equation without additional restrictions or very strong prior information.

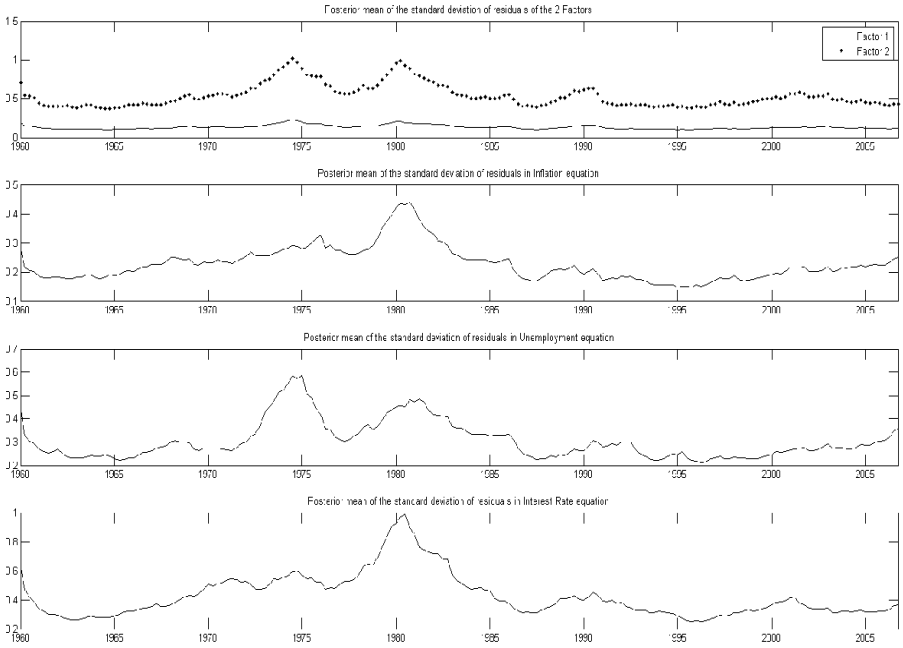

Figure 8: Time-varying volatilities of errors in five key equations of the TVP-FAVAR

Figures 9 and 10 plot impulse responses in the same format as Figures 6 and 7. With the TVP-FAVAR these will be time varying so we plot them for three different time periods. A comparison of Figure 9 to Figure 4 (which presents the same impulse responses using a TVP-VAR) indicate that broad patterns are roughly similar, but there are some important differences between the two figures. Similarly, a comparison of Figure 10 to Figure 8 indicates broad similarities, but in some specific cases important differences can be found.

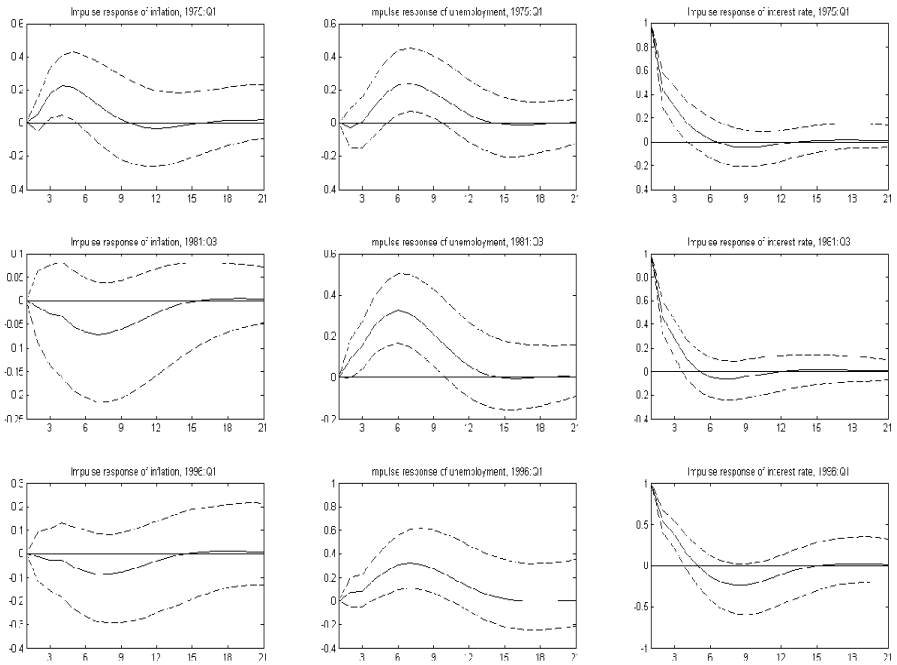

Figure 9: Posterior of impulse responses of main variables to monetary policy shock at different times

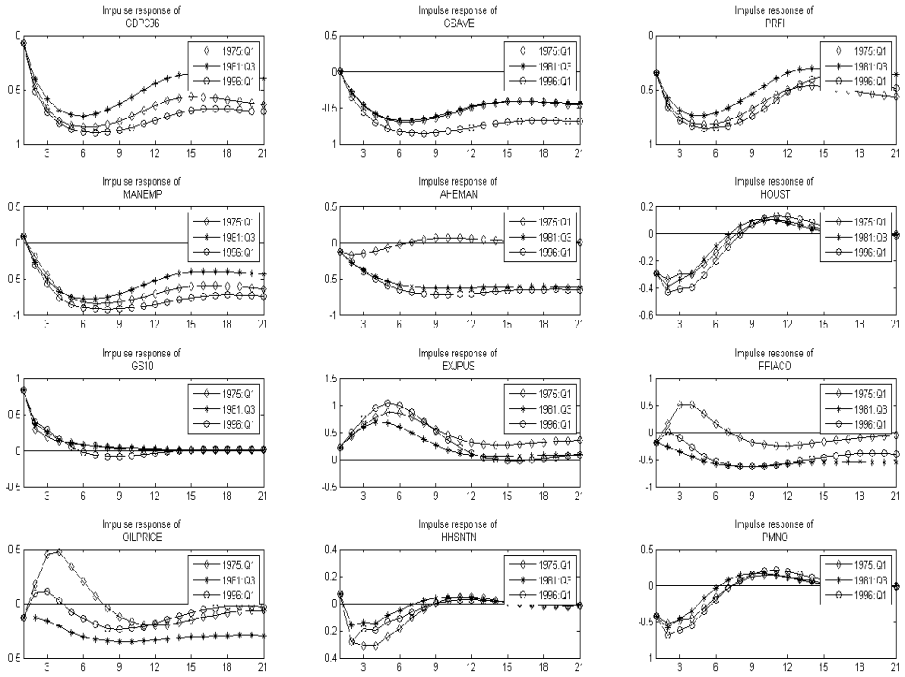

Figure 10: Posterior means impulse responses of selected variables to monetary policy shock at different times

6 Conclusion

In this monograph, we have discussed VARs, TVP-VAR, FAVARs and TVP-FAVARs, including versions of these models with multivariate stochastic volatility. These classes of models have become very popular with empirical macroeconomists since they allow for insight into the relationships between macroeconomic variables in a manner which lets the data speak in a relatively unconstrained manner. However, the cost of working with such unconstrained models is that they risk being over-parameterized. Accordingly, researchers have found it desirable to impose soft or hard restrictions on these models. Soft restrictions typically involve shrinking coefficients towards a particular value (usually zero) whereas hard ones involve the imposition of exact restrictions. Bayesian methods have been found to be an attractive and logically consistent way of handling such restrictions.

We have shown have Bayesian inference can be implemented in a variety of ways in these models, with emphasis on specifications that are of interest to the practitioner. Apart from the simplest of VARs, MCMC algorithms are required. This monograph describes these algorithms in a varying degree of detail. We

draw the reader's attention to the website associated with this monograph which contains Matlab code for most of the models described above. A manual on this website provides a complete listing of all formulae used in each MCMC algorithm. Thus, our aim has been to provide a complete set of Bayesian tools for the practitioner interested in a wide variety of models commonly used in empirical macroeconomics.

References

An, S. and Schorfheide, F. (2007). "Bayesian analysis of DSGE models," *Econometric Reviews*, 26, 113-172.

Asai, M., McAleer, M. and Yu, J. (2006). "Multivariate stochastic volatility: A review," *Econometric Reviews*, 25, 145-175.

Ballabriga, F., Sebastian, M. and Valles, J. (1999). "European asymmetries," *Journal of International Economics*, 48, 233-253.

Banbura, M., Giannone, D. and Reichlin, L. (2008). "Large Bayesian VARs," *Journal of Applied Econometrics*, forthcoming.

Belviso, F. and Milani, F. (2006). "Structural factor augmented VARs (SFAVARs) and the effects of monetary policy," *Topics in Macroeconomics*, 6, 2.

Berg, A., Meyer, R. and Yu, J. (2004). "Deviance information criterion for comparing stochastic volatility models," *Journal of Business and Economic Statistics*, 22, 107-120.

Bernanke, B. and Boivin, J. (2003). "Monetary policy in a data-rich environment," *Journal of Monetary Economics*, 50, 525-546.

Bernanke, B., Boivin, J. and Eliasz, P. (2005). "Measuring monetary policy: A Factor augmented vector autoregressive (FAVAR) approach," *Quarterly Journal of Economics*, 120, 387-422.

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.

Canova, F. (2007). *Methods for Applied Macroeconomic Research* published by Princeton University Press.

Canova, F. (1993). "Modelling and forecasting exchange rates using a Bayesian time varying coefficient model," *Journal of Economic Dynamics and Control*, 17, 233-262.

Canova, F. and Ciccarelli, M. (2009). "Estimating multi-country VAR models," *International Economic Review*, 50, 929-959.

Canova, F. and Ciccarelli, M. (2004). "Forecasting and turning point predictions in a Bayesian panel VAR model," *Journal of Econometrics*, 120, 327-359.

Canova, F. and Gambetti, L. (2009). "Structural changes in the US economy: Is there a role for monetary policy?" *Journal of Economic Dynamics and Control*, 33, 477-490.

Carter, C. and Kohn, R. (1994). "On Gibbs sampling for state space models," *Biometrika*, 81, 541–553.

Chib, S. and Greenberg, E. (1994). "Bayes inference in regression models with ARMA(p,q) errors," *Journal of Econometrics*, 64, 183-206.

Chib, S. and Greenberg, E. (1995). "Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models," *Journal of Econometrics*, 68, 339-360.

Chib, S., Nardari, F. and Shephard, N. (2002). "Markov chain Monte Carlo methods for stochastic volatility models," *Journal of Econometrics*, 108, 281-316.

Chib, S., Nardari, F. and Shephard, N. (2006). "Analysis of high dimensional multivariate stochastic volatility models," *Journal of Econometrics*, 134, 341-371.

Chib, S., Omori, Y. and Asai, M. (2009). "Multivariate stochastic volatility," pages 365-400 in *Handbook of Financial Time Series*, edited by T. Andersen, R. David, J. Kreiss and T. Mikosch. Berlin: Springer-Verlag.

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

Ciccarelli, M. and Rebucci, A., (2002). "The transmission mechanism of European monetary policy: Is there heterogeneity? Is it changing over time?," International Monetary Fund working paper, WP 02/54.

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.

DeJong, P. and Shephard, N. (1995). "The simulation smoother for time series models," *Biometrika*, 82, 339-350.

Del Negro, M. and Otrok, C. (2008). "Dynamic factor models with time varying parameters: Measuring changes in international business cycles," Federal Reserve Bank of New York Staff Report no. 326.

Del Negro, M. and Schorfheide, F. (2004). "Priors from general equilibrium models for VARs," *International Economic Review*, 45, 643-673.

Del Negro, M. and Schorfheide, F. (2009). "Bayesian macroeconometrics," to appear in the *Handbook of Bayesian Econometrics*, edited by J. Geweke, G. Koop and H. van Dijk. Oxford: Oxford University Press.

Doan, T., Litterman, R. and Sims, C. (1984). "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews*, 3, 1-144.

Doucet, A., Godsill, S. and Andrieu, C. (2000). "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, 10, 197-208.

Durbin, J. and Koopman, S. (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

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

Fernandes-Villaverde, J. (2009). "The econometrics of DSGE models," Penn Institute for Economic Research working paper 09-008.

Forni, M. and Reichlin, L. (1998). "Let's get real: A factor analytic approach to disaggregated business cycle dynamics," *Review of Economic Studies*, 65, 453-473.

Fruhwirth-Schnatter, S. (1994). "Data augmentation and dynamic linear models," *Journal of Time Series Analysis*, 15, 183–202.

Fruhwirth-Schnatter, S. and Wagner, H. (2008). "Marginal likelihoods for non-Gaussian models using auxiliary mixture sampling," *Computational Statistics and Data Analysis*, 52, 4608-4624.

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. (2008). "Bayesian stochastic search for VAR model restrictions," *Journal of Econometrics*, 142, 553-580.

Geweke, J. (1977). "The dynamic factor analysis of economic time series," in *Latent Variables in Socio-economic Models*, edited by D. Aigner and A. Goldberger, Amsterdam: North Holland.

Geweke, J. and Keane, M. (2007). "Smoothly mixing regressions," *Journal of Econometrics*, 138, 252-291.

Geweke, J. and Zhou, G. (1996). "Measuring the pricing error of the arbitrage pricing theory," *Review of Financial Studies*, 9, 557-587.

Giordani, P. and Kohn, R. (2008) "Efficient Bayesian inference for multiple change-point and mixture innovation models," *Journal of Business and Economic Statistics*, 26, 66-77.

Giordani, P., Kohn, R. and Pitt, M. (2009). "Time series state space models," to appear in the *Handbook of Bayesian Econometrics*, edited by J. Geweke, G. Koop and H. van Dijk. Oxford: Oxford University Press.

Giordani, P., Kohn, R. and van Dijk, D. (2007). "A unified approach to nonlinearity, structural change and outliers," *Journal of Econometrics*, 137, 112-133.

Groen, J., Paap, R. and Ravazzolo, F. (2008). "Real-time inflation forecasting in a changing world," Erasmus University manuscript, 2008.

Harvey, A. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press: Cambridge.

Ingram, B. and Whiteman, C. (1994). "Supplanting the Minnesota prior - Forecasting macroeconomic time series using real business cycle model priors," *Journal of Monetary Economics*, 49, 1131-1159.

Jacquier, E., Polson, N. and Rossi, P. (1994). "Bayesian analysis of stochastic volatility," *Journal of Business and Economic Statistics*, 12, 371-417.

Johannes, M. and Polson, N. (2009). "Particle filtering," pages 1015-1030 in *Handbook of Financial Time Series*, edited by T. Andersen, R. David, J. Kreiss and T. Mikosch. Berlin: Springer-Verlag.

Kadiyala, K. and Karlsson, S. (1997). "Numerical methods for estimation and inference in Bayesian VAR models," *Journal of Applied Econometrics*, 12, 99-132.

Kim, C. and Nelson, C. (1999). *State Space Models with Regime Switching*. Cambridge: MIT Press.

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

Koop, G. (1992). "Aggregate shocks and macroeconomic fluctuations: A Bayesian approach," *Journal of Applied Econometrics*, 7, 395-411.

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

Koop, G., Leon-Gonzalez, R. and Strachan, R. (2009). "On the evolution of the monetary policy transmission mechanism," *Journal of Economic Dynamics and Control*, 33, 997–1017.

Koop, G., Poirier, D. and Tobias, J. (2007). *Bayesian Econometric Methods*. Cambridge: Cambridge University Press.

Koop, G. and Potter, S. (2004). "Forecasting in dynamic factor models using Bayesian model averaging," *The Econometrics Journal*, 7, 550-565.

Koop, G. and Potter, S. (2006). "The Vector floor and ceiling model," Chapter 4 in *Nonlinear Time Series Analysis of the Business Cycle*, edited by C. Milas, P. Rothman and D. van Dijk in Elsevier's Contributions to Economic Analysis series.

Koop, G. and Potter, S. (2009). "Time varying VARs with inequality restrictions," manuscript available at http://personal.strath.ac.uk/gary.koop/koop_potter14.pdf.

Korobilis, D. (2009a). "Assessing the transmission of monetary policy shocks using dynamic factor models," University of Strathclyde, Discussion Papers in Economics, No. 09-14.

Korobilis, D. (2009b). "VAR forecasting using Bayesian variable selection," manuscript.

Kose, A., Otrok, C. and Whiteman, C. (2003). "International business cycles: World, region and country-specific factors," *American Economic Review*, 93, 1216-1239.

Kuo, L. and B. Mallick. (1997). "Variable selection for regression models," *Shankya: The Indian Journal of Statistics* (Series B), 60, 65-81.

Litterman, R. (1986). "Forecasting with Bayesian vector autoregressions – Five years of experience," *Journal of Business and Economic Statistics*, 4, 25-38.

Lopes, H. and West, M. (2004). "Bayesian model assessment in factor analysis," *Statistica Sinica*, 14, 41-67.

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

Omori, Y., Chib, S., Shephard, N. and Nakajima, J. (2007). "Stochastic volatility with leverage: Fast and efficient likelihood inference," *Journal of Econometrics*, 140, 425-449.

Otrok, C. and Whiteman, C. (1998). "Bayesian leading indicators: Measuring and predicting economic conditions in Iowa," *International Economic Review*, 39, 997-1014.

Paap, R. and van Dijk, H. (2003). "Bayes estimates of Markov trends in possibly cointegrated series: An application to US consumption and income," *Journal of Business and Economic Statistics*, 21, 547-563.

Pitt, M. and Shephard, N. (1999). "Time varying covariances: a factor stochastic volatility approach," pages 547–570 in *Bayesian Statistics*, Volume 6, edited by J.M Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, Oxford: Oxford University Press.

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

Sentana, E. and Fiorentini, G. (2001). "Identification, estimation and testing of conditionally heteroskedastic factor models," *Journal of Econometrics*, 102,

143-164.

Sims, C. (1993). "A nine variable probabilistic macroeconomic forecasting model," pp. 179-204 in J. Stock and M. Watson, eds., *Business Cycles Indicators and Forecasting*. University of Chicago Press for the NBER).

Sims, C. (1980). "Macroeconomics and reality," *Econometrica*, 48, 1–48.

Sims, C. and Zha, T. (1998). "Bayesian methods for dynamic multivariate models," *International Economic Review*, 39, 949-968.

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

Stock, J. and Watson, M., (1996). "Evidence on Structural Instability in Macroeconomic Time Series Relations," *Journal of Business and Economic Statistics,* 14, 11-30.

Stock, J. and Watson, M., (1999). "Forecasting inflation," *Journal of Monetary Economics,* 44, 293-335.

Stock, J. and Watson, M., (2002). "Macroeconomic forecasting using diffusion indexes," *Journal of Business and Economic Statistics,* 20, 147-162.

Stock, J. and Watson, M., (2005). "Implications of dynamic factor models for VAR analysis," National Bureau of Economic Research working paper 11467.

Stock, J. and Watson, M., (2006). "Forecasting using many predictors," pp. 515-554 in *Handbook of Economic Forecasting*, Volume 1, edited by G. Elliott, C. Granger and A. Timmerman, Amsterdam: North Holland.

Villani, M. (2009). "Steady-state prior for vector autoregressions," *Journal of Applied Econometrics*, 24, 630-650.

West, M. (2003). "Bayesian factor regression models in the 'large p, small n' paradigm," pages 723-732 in *Bayesian Statistics*, Volume 7, edited by J.M. Bernardo, M. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, and M. West, Oxford: Oxford University Press.

West, M. and Harrison, P. (1997). *Bayesian Forecasting and Dynamic Models*. Second edition. Berlin: Springer.