Time Varying VARs with Inequality Restrictions[*]

Gary Koop[†]

Department of Economics

University of Strathclyde

email: Gary.Koop@strath.ac.uk

Simon M. Potter

Macroeconomics and Monetary Studies Function

Federal Reserve Bank of New York

email: simon.potter@ny.frb.org

November 2008

Abstract

In many applications involving time-varying parameter VARs, it is desirable to restrict the VAR coefficients at each point in time to be non-explosive. This is an example of a problem where inequality restrictions are imposed on states in a state space model. In this paper, we describe how existing MCMC algorithms for imposing such inequality restrictions can work poorly (or not at all) and suggest alternative algorithms which exhibit better performance. Furthermore, previous algorithms involve an approximation relating to a key integrating constant. Our algorithms are exact, not involving this approximation. In an application involving a commonly-used U.S. data set, we show how

---

this approximation can be a poor one and present evidence that the algorithms proposed in this paper work well.

1 Introduction

Time varying parameter (TVP) models have a long history in economics (e.g. Cooley and Prescott, 1973, 1976). Recently, they have been influential in empirical macroeconomics in papers such as (among many others) Cogley and Sargent (2001, 2005), Kim and Nelson (1999), Primiceri (2005) and Stock and Watson (2007). Such papers use (sometimes restricted variants) of state space models involving a measurement equation:

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

and a state equation

$$\theta_t = \theta_{t-1} + \eta_t, \tag{2}$$

where y_t is an $p \times 1$ vector of observations on the dependent variables, θ_t an $m \times 1$ vector of unobserved states, ε_t are independent $N(0, H_t)$ random vectors and η_t are independent $N(0, Q)$ random vectors for $t = 1, .., T$.[1] The errors in the two equations, ε_t and η_s, are independent of one another for all t and s.[2] Z_t is the appropriate $p \times m$ matrix of data on explanatory variables. For instance, Cogley and Sargent (2001, 2005) and Primiceri (2005) use TVP-VARs and, hence, each row of Z_t contains lags of all dependent variables and an intercept and other deterministic terms and θ_t are the VAR coefficients. Stock and Watson (2007) use a model where y_t is inflation, $Z_t = 1$ and θ_t is interpreted as underlying inflation.

Econometric inference in the model defined by (1) and (2) can be carried out using standard state space methods involving the Kalman filter. For instance, Carter and Kohn (1994), DeJong and Shephard (1995) and Durbin and Koopman (2002) provide algorithms for drawing $\theta^T = (\theta_1', .., \theta_T')'$ which are commonly used in Bayesian empirical work as part of a Markov Chain

---

[1] Stock and Watson (2007) allow for Q to vary over time as well. To allow for this could be done using extensions of the methods developed in this paper.

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

Monte Carlo (MCMC) algorithm. However, these methods only work when θ^T is unrestricted (i.e. it is not subject to any inequality restrictions). In many macroeconomic applications, we wish to impose inequality restrictions on the states. For instance, the model of Stock and Watson (2007) uses an unrestricted state space model which implies that underlying inflation follows a random walk and, thus, can grow in an unbounded fashion. If central banks have implicit or explicit inflation targets, this may be unreasonable and the researcher may wish to limit underlying inflation to lie within bounds (see Koop and Potter, 2008). As another example, in TVP-VAR models it can be desirable to impose stability on the TVP-VAR at each point in time. This leads papers such as Cogley and Sargent (2001, 2005) to restrict θ_t to be non-explosive. This involves the inequality restriction that the roots of the VAR polynomial defined by θ_t lie outside the unit circle.[3] Indeed, in the absence of such inequality restrictions (or a very tight prior), Bayesian TVP-VARs will place a large amount of a priori weight on nonsensical paths for the states. This can cause substantive problems for empirical work. For instance, even a small amount of posterior weight in explosive regions for θ_t can lead to impulse responses or forecasts which have counter-intuitively huge posterior means or standard deviations.

Such considerations motivate why imposing inequality restrictions on the states in state space models can be very important. However, with the one exception which we will discuss below (i.e. the algorithm described in Cogley and Sargent, 2005), to our knowledge there do not exist any MCMC algorithms for drawing from state space models subject to inequality restrictions[4].The purpose of this paper is to fill this gap. We develop several MCMC algorithms and investigate their performance in the TVP-VAR used in Primiceri (2005).

To informally motivate our algorithms, note that they improve on two possible strategies. First, the naive researcher may think that a standard algorithm such as that of Durbin and Koopman (2002) can be used, with the added restriction that, if a drawn θ_t violates the inequality restriction, then it can be immediately discarded and new $\theta_t s$ can be drawn until one

---

[3]Note, though, that Cogley and Sargent (2001, 2005) do not restrict the intercepts. Thus, the intercepts have a random walk structure and the elements of y_t have unit roots. To rule out unit root behavior, one can further restrict the intercepts to bounded intervals.

[4]See Koop, Leon-Gonzalez and Strachan (2008) for a Bayesian analaysis of state space models where the states are subject to *equality* restrictions at some time periods but not others.

satisfies the restriction. Such a strategy is simply wrong: it does not lead to a valid posterior simulator. A second strategy, adopted by Cogley and Sargent (2005), is to use an algorithm such as Carter and Kohn (1994) or Durbin and Koopman (2002) to draw an entire vector of states, $\theta^T = (\theta_1', .., \theta_T')'$, and then discard this entire vector of states if any of the $\theta_t s$ violate the constraints. In this paper, we elaborate why their strategy also runs into problems. One of these problems is that it is very easy to run into cases where virtually all of the draws are discarded and computation is not feasible in a reasonable amount of time. For instance, in an unrestricted TVP-VAR, even in the very well-behaved case where all of the $\theta_t s$ have posterior medians which are non-explosive, it is common for posterior standard deviations to be fairly large (and, thus, some posterior probability is attached to explosive regions of the parameter space). If each of T states has a small chance of violating the restrictions, it is not hard to get the probability that at least one of them will violate the restrictions being very nearly one in every draw. In such cases, the algorithm of Cogley and Sargent (2005) will discard nearly every draw. It is not hard to find empirical problems where such an algorithm can take billions and billions of draws without accepting even a single one.

In this paper, we develop improvements on these two sorts of strategies. We begin with the development of a multi-move algorithm similar to that of Cogley and Sargent (2005), but not involving a certain approximation which they use. We recommend this algorithm for cases where the unrestricted posterior indicates high probability in favor of the restrictions holding for all of the states. We also develop a single-move algorithm which draws the states one at a time. This algorithm is recommended for cases where the multi-move algorithm discards virtually every candidate draw. However, because it is a single-move algorithm it can mix more slowly.

The remainder of the paper is organized as follows. In the next section, we derive our MCMC algorithms. In the third section, we present empirical evidence relating to their performance. The fourth section concludes.

2   MCMC Algorithms for State Space Models Subject to Inequality Restrictions

In this section, we derive several MCMC algorithms for drawing from the state space model defined by (1) and (2) subject to the inequality restrictions

that $\theta_t \in A$ for some region A. For instance, in our empirical work we focus on the TVP-VAR case and the restriction is that the roots of the matrix polynomial defined by the VAR coefficients in θ_t lie outside the unit circle. The contributions of this paper relate to drawing θ_t and Q. Thus, to keep the notation as uncluttered as possible, we do not include H_t as a conditioning argument in the posterior conditional distributions given below. It is common to specify a multivariate stochastic volatility form for H_t and, in our empirical work, we do so (see the Appendix for details). However, this merely adds more blocks to our MCMC algorithm of standard form, so we do not discuss it in this section. Also, we abstract from treatment of the initial condition for the states, θ_0 and just assume a standard treatment is used (and do not list θ_0 as a conditioning argument in the probabilities below). For instance, if a regression effect (i.e. a term with constant coefficient, $Z_t \widetilde{\theta}$ is added to the measurement equation), then we have $\theta_0 = 0$. Alternatively, as in Durbin and Koopman, a Normal prior such as $\theta_0 \sim N(\underline{\theta}_0, \underline{V}_0)$ can be used for initial condition. Thus, we focus on the states $\theta^T = (\theta'_1, .., \theta'_T)'$ and the error covariance matrix in the state equation Q. We use the notation that $y^T = (y'_1, .., y'_T)'$ is the entire sample.

There are some general results that are used by some or all of the MCMC algorithms we develop. We describe these here. Our multi-move algorithms are Metropolis-Hastings algorithms which take candidate draws from the posterior for the model defined by (1) and (2) without the inequality restrictions. We refer to this model as being "unrestricted" and place "U" subscripts on densities relating to this model. Standard algorithms (e.g. Durbin and Koopman, 2002) can be used to provide MCMC draws from the unrestricted model.

All of the MCMC algorithms below involve draws from the unrestricted posterior of Q, conditional on the states, $p_U\left(Q|\theta^T, y^T\right)$. This can also be done using standard methods. For instance, if we follow the common practice of using a Wishart prior: $Q^{-1} \sim W\left(\underline{\nu}_Q, \underline{Q}^{-1}\right)$, then the unrestricted posterior for Q^{-1} (conditional on the states) is Wishart:

$$Q^{-1}|\theta^T, y^T \sim W\left(\overline{\nu}_Q, \overline{Q}^{-1}\right) \tag{3}$$

where

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

and

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

In our empirical work, we use a training sample to choose the prior hyper-parameters (see Appendix for details).

2.1 Multi-move algorithms

We begin by describing two multi-move Metropolis-Hastings algorithms for drawing the unobserved states. That is, in these algorithms, the previous draw of the sequence of states is not used directly in the construction of the new draw. An alternative is to use a single move sampler that generates the draw of θ_t conditional on θ_{-t} where $\theta_{-t} = \left(\theta_1', .., \theta_{t-1}', \theta_{t+1}', .., \theta_T'\right)'$. Usually multi-move samplers are preferred because of the reduction in dependence across draws of the chain they produce. But, in the present case, a single move sampler might have certain advantages since (for reasons discussed previously) the multi-move algorithm could lead to extremely low Metropolis-Hastings acceptance probabilities (i.e. where virtually all candidate draws are rejected).

The posterior of interest is $p\left(\theta^T, Q|y^T\right)$, with prior $p\left(\theta^T, Q\right)$ and likelihood $p\left(y^T|\theta^T, Q\right)$. Given the structure of the model, the likelihood can be written as:

$$p\left(y^T|\theta^T, Q\right) = \prod_{t=1}^{T} p(y_t|\theta_t, Q), \tag{4}$$

where, for notational simplicity, we are not explicitly listing Z_t as a conditioning argument. Note that the likelihood for the unrestricted posterior also has this form so that $p_U\left(y^T|\theta^T, Q\right) = p\left(y^T|\theta^T, Q\right)$. Furthermore, we will also use the same prior for Q so that $p_U\left(Q\right) = p\left(Q\right)$.

The prior is given by:

$$
\begin{aligned}
p\left(\theta^T, Q\right) &= p\left(\theta^T|Q\right) p\left(Q\right) \\
&\propto \left\{ \prod_{t=1}^{T} 1(\theta_t \in A) p_U(\theta_t|\theta_{t-1}, Q) \right\} p\left(Q\right),
\end{aligned}
\tag{5}
$$

6

where $p_U(\theta_t|\theta_{t-1}, Q)$ is given by (2), without the restrictions imposed, and $1(.)$ is the indicator function. Multiplying prior by likelihood, we obtain the posterior:

$$p\left(\theta^T, Q|y^T\right) = \frac{\left\{\prod_{t=1}^{T} p(\theta_t|\theta_{t-1}, Q)p(y_t|\theta_t, Q)\right\} p\left(Q\right)}{m\left(y^T\right)}, \tag{6}$$

where $m\left(y^T\right)$ is the marginal likelihood and $p(\theta_t|\theta_{t-1}, Q)$ is obtained from the state equation with the restrictions imposed:

$$p(\theta_t|\theta_{t-1}, Q) \propto \frac{1(\theta_t \in A)}{2\pi^{\frac{m}{2}}} |Q|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\theta_t - \theta_{t-1}\right)' Q^{-1}\left(\theta_t - \theta_{t-1}\right)\right],$$

and we denote the integrating constant for this prior by $R\left(\theta_{t-1}, Q\right)$. Note that we have a relationship between our prior and the comparable prior for the unrestricted model

$$p(\theta_t|\theta_{t-1}, Q) = \frac{1(\theta_t \in A)p_U(\theta_t|\theta_{t-1}, Q)}{R\left(\theta_{t-1}, Q\right)}, \tag{7}$$

for $t = 1, .., T$.

If we use the same treatment of the initial condition in the unrestricted and restricted model, the unrestricted posterior is exactly the same as (6) except that the $\frac{1(\theta_{tt}\in A)}{R(\theta_{t-1}, Q)}$ terms are omitted and $m\left(y^T\right)$ is replaced by $m_U\left(y^T\right)$ where the latter is the marginal likelihood for the unrestricted posterior.

The standard MCMC algorithm (without inequalities) involves drawing from $p_U\left(\theta^T|Q, y^T\right)$ and $p_U\left(Q|y^T, \theta^T\right)$ and we can use these as candidate generating densities in a Metropolis-Hastings algorithm when dealing with the restricted model. This requires us to calculate the Metropolis-Hastings acceptance probabilities. Consider first the states and let $s = 1, .., S$ index (post burn-in) draws from the algorithm and superscripts (s) denote the accepted draws. Let superscripts $(*)$ denote candidate draws so that $\theta^{T(*)}$ is a candidate draw from $p_U\left(\theta^T|y^T, Q^{(s-1)}\right)$ which is accepted with probability α_θ. The formula for Metropolis-Hasting acceptance probability, using $p_U\left(\theta^T|y^T, Q^{(s-1)}\right)$ as a candidate generating density, is (see, e.g., Koop, 2003, page 94):

$$\alpha_\theta = \min \left[\frac{p\left(\theta^T = \theta^{T(*)}|y^T, Q^{(s-1)}\right) p_U\left(\theta^T = \theta^{T(s-1)}|y^T, Q^{(s-1)}\right)}{p\left(\theta^T = \theta^{T(s-1)}|y^T, Q^{(s-1)}\right) p_U\left(\theta^T = \theta^{T(*)}|y^T, Q^{(s-1)}\right)}, 1 \right]$$

$$= \min \left[\frac{w_\theta(*)}{w_\theta(s-1)}, 1 \right]$$

where

$$w_\theta(*) = \frac{p\left(\theta^T = \theta^{T(*)}|y^T, Q^{(s-1)}\right)}{p_U\left(\theta^T = \theta^{T(*)}|y^T, Q^{(s-1)}\right)}$$

and

$$w_\theta(s-1) = \frac{p\left(\theta^T = \theta^{T(s-1)}|y^T, Q^{(s-1)}\right)}{p_U\left(\theta^T = \theta^{T(s-1)}|y^T, Q^{(s-1)}\right)}.$$

Note that, since $p\left(\theta^T|y^T, Q\right) = p\left(\theta^T, Q|y^T\right)/p\left(Q|y^T\right)$ (and the same holds for the unrestricted model), we can just plug in the formula for $p\left(\theta^T, Q|y^T\right)$ instead of $p\left(\theta^T|y^T, Q\right)$ in the formula for the acceptance probability (and the two terms $p\left(Q|y^T\right)$ will cancel out as will the two $p_U\left(Q|y^T\right)$ terms). Thus, we can use our previous derivations to say:

$$w_\theta(*) = \frac{m_U\left(y^T\right) \prod_{t=1}^{T} \frac{1(\theta_t^{(*)} \in A) p_U(\theta_t^{(*)}|\theta_{t-1}^{(*)}, Q^{(s-1)})}{R\left(\theta_{t-1}^{(*)}, Q^{(s-1)}\right)} p(y_t|\theta_t^{(*)}, Q^{(s-1)})}{m\left(y^T\right) \prod_{t=1}^{T} p_U(\theta_t^{(*)}|\theta_{t-1}^{(*)}, Q^{(s-1)}) p_U(y_t|\theta_t^{(*)}, Q^{(s-1)})}$$

$$= \frac{m_U\left(y^T\right)}{m\left(y^T\right)} \prod_{t=1}^{T} \frac{1(\theta_t^* \in A)}{R\left(\theta_{t-1}^{(*)}, Q^{(s-1)}\right)}.$$

An identical derivation, using the additional fact that $1(\theta_t^{(s-1)} \in A) = 1$ (since the previous draw was an accepted one, it must be the case that it satisfies the inequality restriction), gives

$$w_\theta(s-1) = \frac{m_U\left(y^T\right)}{m\left(y^T\right)} \prod_{t=1}^{T} \frac{1}{R\left(\theta_{t-1}^{(s-1)}, Q^{(s-1)}\right)}$$

8

and, thus,

$$\alpha_\theta = \min \left[\prod_{t=1}^{T} \frac{\frac{1(\theta_t^{(*)} \in A)}{R\left(\theta_{t-1}^{(*)}, Q^{(s-1)}\right)}}{\frac{1}{R\left(\theta_{t-1}^{(s-1)}, Q^{(s-1)}\right)}}, 1 \right]. \tag{8}$$

Now consider drawing Q in the Metropolis-Hastings algorithm after drawing θ^T (i.e. conditional on $\theta^{T(s)}$). Derivations are almost the same as for θ^T. A Metropolis-Hastings algorithm takes candidate draws from $p_U \left(Q | y^T, \theta^{T(s)}\right)$ and these are accepted with probability

$$\alpha_Q = \min \left[\prod_{t=1}^{T} \frac{R(\theta_{t-1}^{(s)}, Q^{(s-1)})}{R(\theta_{t-1}^{(s)}, Q^{(*)})}, 1 \right], \tag{9}$$

where the derivations are as above, but we are using the fact that $1(\theta_t^{(s)} \in A) = 1$ for all t.

Note also that $\frac{m(y^T)}{m_U(y^T)}$ is the Bayes factor comparing the restricted to unrestricted model. For future reference, it is worth noting that it can be calculated using MCMC output from the unrestricted posterior and the relationship:

$$BF = \frac{m\left(y^T\right)}{m_U\left(y^T\right)} = E \left[\Pi_{t=1}^T \frac{1(\theta_t \in A)}{R(\theta_{t-1}, Q)} \right], \tag{10}$$

where the expectation is taken with respect to the unrestricted posterior. That is, we can derive:

$$
\begin{aligned}
& E \left[\Pi_{t=1}^T \frac{1(\theta_t \in A)}{R(\theta_{t-1}, Q)} \right] \\
=\ & \int \Pi_{t=1}^T \frac{1(\theta_t \in A)}{R(\theta_{t-1}, Q)} p_U \left(\theta^T, Q | y^T\right) d\theta^T dQ \\
=\ & \int \Pi_{t=1}^T \frac{1(\theta_t \in A)}{R(\theta_{t-1}, Q)} \frac{\left\{\Pi_{t=1}^T p_U(\theta_t | \theta_{t-1}, Q) p(y_t | \theta_t, Q)\right\} p(Q)}{m_U\left(y^T\right)} d\theta^T dQ \\
=\ & \frac{1}{m_U\left(y^T\right)} \int \Pi_{t=1}^T \frac{1(\theta_t \in A)}{R(\theta_{t-1}, Q)} \left\{ \prod_{t=1}^T p_U(\theta_t | \theta_{t-1}, Q) p(y_t | \theta_t, Q) \right\} p(Q) d\theta^T dQ \\
=\ & \frac{m\left(y^T\right)}{m_U\left(y^T\right)}.
\end{aligned}
$$

9

Thus, MCMC output from the unrestricted posterior can be used to calculate this Bayes factor. It should be emphasized that this requires a separate run of the MCMC algorithm beyond that required to carry out posterior inference in the model. That is, the Metropolis-Hastings algorithm described above provides a chain of draws from the restricted posterior, not the unrestricted posterior. To calculate the Bayes factor we need a chain of draws from the unrestricted posterior. To be precise, our Metropolis-Hasting algorithm takes candidate draws from $p\left(\theta^T|y^T, Q^{(s-1)}\right)$ where the $Q^{(s-1)}$ is a draw from the restricted posterior. But to calculate the Bayes factor requires draws from $p\left(\theta^T|y^T, Q^{(s-1)}\right)$ where the $Q^{(s-1)}$ is a draw from the unrestricted posterior.

To aid in interpretation of the Bayes factor, note that if all the draws from the unrestricted posterior satisfy the restrictions then the Bayes factor will favor the restricted model (since $R(\theta_{t-1}, Q) \leq 1$). However, if very few of the draws satisfy the restrictions then the unrestricted model is favored unless $\prod_{t=1}^{T} R(\theta_{t-1}, Q)$ is very low for the draws which satisfy the restrictions. This can be seen in another way by noting that (10) can also be written as

$$BF = \Pr\left[\text{Restrictions are true}\right] E\left[\frac{1}{\prod_{t=1}^{T} R(\theta_{t-1}, Q)}|\text{Restrictions are true}\right]$$
(11)

where $\Pr\left[\text{Restrictions are true}\right]$ is the proportion of MCMC draws from the unrestricted posterior which satisfy the restrictions. Note first that a naive researcher might be tempted to use $\Pr\left[\text{Restrictions are true}\right]$ alone as a measure of support for the restrictions, but (11) makes clear that this is not enough. In practice, especially if $\Pr\left[\text{Restrictions are true}\right]$ is low the role of the expected value term in (11) can be quite substantial. Since $\prod_{t=1}^{T} R(\theta_{t-1}, Q) \leq 1$ (and often much less than one), the effect of ignoring this term would be to under-report the support for the restrictions.

Thus, the unrestricted posterior can be used as a candidate generating density in a Metropolis-Hastings algorithm with acceptance probabilities given by (8) and (9). The efficiency of this approach will depend on the probability that the restrictions are true under the unrestricted posterior. Such a strategy is very similar to what is done in Cogley and Sargent (2005). However, they use an approximation, with $\alpha_Q = 1$ and an acceptance probability

for the states of:

$$\widehat{\alpha}_\theta = \prod_{t=1}^{T} 1(\theta_t \in A).$$ (12)

In words, they use standard state space methods to draw a vector θ^T from the unrestricted posterior and then attach an acceptance probability of zero if any drawn θ_t violates the restrictions, otherwise they accept the draw. Their justification is implicitly based on the assumption that $R(\theta_{t-1}, Q) \approx c$ (for some constant c) for all values of $\theta_{t-1} \in A$ and Q. This is likely to be the case if most of the draws satisfy the restrictions and $R(\theta_{t-1}, Q) \approx 1$. However, if few draws do, then it is possible that the approximation of Cogley and Sargent is quite poor. This is an issue we investigate empirically below.

A problem with these multi-move algorithms is that they both depend on the term $\prod_{t=1}^{T} 1(\theta_t \in A)$. This will be equal to zero even if only a single θ_t violates the restriction. Thus, it is possible (and often happens in practice) that Metropolis-Hastings acceptance probabilities will be zero for virtually every draw. This motivates consideration of a single-move sampler which, although it may be slow to mix, will yield more reasonable Metropolis-Hastings acceptance probabilities.

2.1.1 Evaluating the Probability of the Restriction Set

The approximation used by Cogley and Sargent is a useful one in that $R(\theta_{t-1}, Q)$ does not have to be evaluated. But our multi-move Metropolis-Hastings algorithm requires a method to evaluate $R(\theta_{t-1}, Q)$ In a few cases it will be possible to analytically evaluate this. Consider, for instance, the case where a scalar θ_t is restricted to lie in the interval $[a, b]$. In this case,

$$R(\theta_{t-1}, Q) = \Phi\left(\frac{b - \theta_{t-1}}{\sqrt{Q}}\right) - \Phi\left(\frac{a - \theta_{t-1}}{\sqrt{Q}}\right)$$

and Φ is the cumulative distribution of the standard Normal.

But, in general, the evaluation of $R(\theta_{t-1}, Q)$ will require simulation methods. To some extent, the degree of computational difficulty depends on whether A consists of linear restrictions or whether it includes some nonlinear restrictions. For linear restrictions, methods such as the GHK simulator will allow for fairly quick calculation of $R(\theta_{t-1}, Q)$ (see, e.g., Geweke, 1991). However, for nonlinear restrictions, there is no general alternative but to

generate draws from the appropriate unrestricted multivariate Normal distribution and count the proportion of times the restriction is satisfied. This is very simple to implement, but can be slow since it must be done for every one of the T states (both when calculating α_θ and α_Q). Nevertheless, for our TVP-VAR applications, we do not find such a strategy places an insuperable computational burden. As a rough order of magnitude calculation, if we (at each MCMC replication for each of $t = 1,..,T$ draws of θ_t) simulate 500 draws to approximate $R(\theta_{t-1}, Q)$ it multiplies the time taken by our MCMC algorithm about tenfold.

However, the use of simulation methods to calculate Metropolis-Hastings acceptance probabilities raises a minor issue which arises since this probability is bounded at one. Let $j = 1,..,J$ index the simulated values used to approximate each of the $R(\theta_{t-1}, Q)$ which appear in

$$\widetilde{\alpha}_\theta = \prod_{t=1}^{T} \frac{\frac{1(\theta_t^{(*)} \in A)}{R\left(\theta_{t-1}^{(*)}, Q^{(s-1)}\right)}}{\frac{1}{R\left(\theta_{t-1}^{(s-1)}, Q^{(s-1)}\right)}}.$$

If J is large, we can use asymptotic results to argue that the following result holds approximately:

$$\widetilde{\alpha}_\theta \sim N\left(\overline{\widetilde{\alpha}}_\theta, \overline{v}_\theta\right),$$

where

$$\overline{\widetilde{\alpha}}_\theta = \frac{\sum_{j=1}^{J} \widetilde{\alpha}_\theta^{(j)}}{J}$$

and

$$\overline{v}_\theta = \frac{\sum_{j=1}^{J} \left(\widetilde{\alpha}_\theta^{(j)} - \overline{\widetilde{\alpha}}_\theta\right)^2}{J}.$$

However, we are not directly interested in $\widetilde{\alpha}_\theta$, but

$$\alpha_\theta = \min\left[\widetilde{\alpha}_\theta, 1\right].$$

Thus, using $\min\left[\overline{\widetilde{\alpha}}_\theta, 1\right]$ would lead to a biased estimate of α_θ. If $\overline{v}_\theta$ is very small (i.e. enough draws are taken to approximate $R(\theta_{t-1}, Q)$ to a high degree of accuracy), then this bias will be empirically negligible. However, to be on

the safe side, we use the formula for the mean of the truncated Normal distribution to obtain an estimate of α_θ. Adding the fact that probability is truncated at zero, we use as our estimate of the acceptance probability:

$$\overline{\alpha}_\theta = \overline{\overline{\alpha}}_\theta + \sqrt{\overline{v}_\theta} \frac{\phi\left(\frac{-\overline{\overline{\alpha}}_\theta}{\sqrt{\overline{v}_\theta}}\right) - \phi\left(\frac{1-\overline{\overline{\alpha}}_\theta}{\sqrt{\overline{v}_\theta}}\right)}{\Phi\left(\frac{1-\overline{\overline{\alpha}}_\theta}{\sqrt{\overline{v}_\theta}}\right) - \Phi\left(\frac{-\overline{\overline{\alpha}}_\theta}{\sqrt{\overline{v}_\theta}}\right)},$$

where $\phi\left(\right)$ is the p.d.f. of the standard Normal distribution. Similar considerations hold for the acceptance probability in the algorithm for drawing Q and a similar transformation is done.

2.2   A single move algorithm

We maintain the same setup, notation and assumptions as before. Our method for drawing Q is exactly the same as the multi-move algorithm and, hence, in this section we focus only on the conditional posterior $p\left(\theta^T|y^T, Q\right)$. The multi-move algorithms discussed previously draw from this directly (or approximately in the case of the Cogley-Sargent algorithm). Our single move algorithm involves drawing sequentially from $p\left(\theta_t|, y^T, \theta_{-t}, Q\right)$. The proposal density for our Metropolis-Hastings algorithm is the unrestricted conditional posterior $p_U\left(\theta_t|, y^T, \theta_{-t}, Q\right)$. By the first order Markov assumption in the state equation, this simplifies to:

$$p_U\left(\theta_t|y^T, \theta_{t-1}, \theta_{t+1}, Q\right) \tag{13}$$

$$\propto \prod_{i=-1,1} \exp\left[-\frac{1}{2}\left(\theta_{t+i} - \theta_t\right)' Q^{-1}\left(\theta_{t+i} - \theta_t\right)\right]$$

$$\times \exp\left[-\frac{1}{2}\left(y_t - Z_t\theta_t\right)' H_t^{-1}\left(y_t - Z_t\theta_t\right)\right]$$

for $t = 1, ..., T-1$.

This is a Normal density with mean vector:

$$\mu_t = \frac{\theta_{t-1} + \theta_{t+1}}{2} + G_t\left[y_t - Z_t\left(\frac{\theta_{t-1} + \theta_{t+1}}{2}\right)\right],$$

where

$$G_t = \frac{1}{2}QZ_t'\left(Z_tQZ_t' + H_t\right)^{-1}$$

13

and covariance matrix:

$$\Omega_t = \frac{1}{2} \left(I_m - G_t Z_t \right) Q.$$

For the final state, the unrestricted posterior that we draw from is $p_U \left(\theta_T | \theta_{T-1}, Q, y^T \right)$ which is also a Normal density, but has mean given by

$$\mu_T = \theta_{T-1} + G_T \left[y_T - Z_T \theta_{T-1} \right],$$

where

$$G_T = Q Z_T' \left(Z_T Q Z_T' + H_T \right)^{-1}$$

and covariance matrix:

$$\Omega_T = \left(I_m - G_T Z_T \right) Q,$$

We wish to draw from the following restricted conditional posterior $p \left(\theta_t | y^T, \theta_{t-1}, \theta_{t+1}, Q \right)$. The derivation basically involves the same steps as for the unrestricted posterior. However, a slight complication arises due to the way the decomposition works and the fact that the integrating constant for the conditional prior for θ_{t+1} depends on θ_t. To see this clearly, remember our previous notation where $R \left(\theta_{t-1}, Q \right)$ is the integrating constant from (7). For $t = 1, .., T - 1$, our restricted conditional posterior has the form:

$$p \left(\theta_t | y^T, \theta_{t-1}, \theta_{t+1}, Q \right) \propto \frac{1(\theta_t \in A)}{R \left(\theta_t, Q \right)} p_U \left(\theta_t | \theta_{t-1}, \theta_{t+1}, Q, y^T \right). \qquad (14)$$

The remaining integrating constant in (14) will cancel out in the acceptance probability and can be ignored.

In a similar fashion as with our multi-move algorithm, we can use the relationship between the unrestricted posterior in (13) and the restricted posterior in (14) in order to obtain an acceptance probability for a candidate draw of θ_t which we will label $\theta_t^{(*)}$. Thus, we have

$$
\begin{aligned}
\alpha_{\theta t} &= \min \left[\frac{\frac{1(\theta_t^{(*)} \in A)}{R \left(\theta_t^{(*)}, Q^{(s-1)} \right)}}{\frac{1}{R \left(\theta_t^{(s-1)}, Q^{(s-1)} \right)}}, 1 \right] \qquad (15) \\
&= \min \left[\frac{1(\theta_t^{(*)} \in A) R \left(\theta_t^{(s-1)}, Q^{(s-1)} \right)}{R \left(\theta_t^{(*)}, Q^{(s-1)} \right)}, 1 \right].
\end{aligned}
$$

14

For $t = T$, this simplifies to

$$\alpha_{\theta T} = 1(\theta_T^{(*)} \in A).$$

We have now described our two sets of algorithms. Each has its advantages and disadvantages. In general, multi-move algorithms will mix faster, but if they end up rejecting virtually every draw then they may not mix at all. Single move algorithms, in general, will mix more slowly, but are less likely to get stuck in this manner. Of course, in practice, there is no need to choose one approach or the other and an algorithm which combines the two could be used (e.g. a multi-move algorithm could be used provided it did not get stuck, if so the algorithm could switch to the single move method).

3   Application to a TVP-VAR

In order to investigate the properties of our algorithms, we use the model of Primiceri (2005) which is similar to that used by Cogley and Sargent (2005) and others. It is a TVP-VAR as specified in (1) and (2) with the additional assumption that the measurement error covariance matrix, H_t, exhibits multivariate stochastic volatility of a particular sort (see the appendix for details). We use data from 1953Q1 through 2006Q2 on three commonly-used variables: the unemployment rate (seasonally adjusted civilian unemployment rate, all workers over age 16), interest rate (yield on three month Treasury bill rate) and inflation rate (the annual percentage change in a chain-weighted GDP price index).[5]

Primiceri (2005) uses a training sample prior involving the first ten years of data. To be precise, he uses the training sample and a time-invariant VAR to produce OLS estimates of the VAR coefficients, $\widehat{\theta}$, and the associated covariance matrix, $\widehat{V}$. For Q, the prior given in (??) is used with $\underline{\nu}_Q = 40$ and $\underline{Q} = 0.0001\widehat{V}$. This is the prior we call the "Primiceri Prior" in our empirical results. We also present results for what we call the "Alternative Prior" which has $\underline{\nu}_Q = 4$ and $\underline{Q} = 0.01\widehat{V}$. This alternative prior is one which could reasonably be entertained by a researcher, but is less informative and centers the prior over a region of the parameter space which allows for slightly

---

[5]The data were obtained from the Federal Reserve Bank of St. Louis website, http://research.stlouisfed.org/fred2/. Note that Primiceri (2005) used data through 2001Q3. Hence, our data is not identical to his.

more time-variation in the VAR coefficients. The prior for the remaining parameters is discussed in the appendix.

Note that we are deliberately using a much more complicated model than is necessary. Our model uses the algorithms developed above as a set of blocks in a bigger MCMC algorithm (e.g. involving sets of blocks for the multivariate stochastic volatility, etc.). We do this deliberately in order to investigate the performance of these algorithms in as realistic a context as possible. Of course, by adding more blocks our algorithms will mix more slowly (and finding out just how slow is an issue we wish to investigate). But, since the additional blocks relating to multivariate stochastic volatility, etc. will be identical in all our algorithms, the relative comparison of multi-move and single move algorithms will be little affected by their inclusion.

We use the inequality restriction that the roots of the matrix polynomial defined by the VAR coefficients in θ_t lie outside the unit circle. We begin with a very brief summary of the performance of the multi-move algorithms using the training sample prior of Primiceri (2005), making his choice of two lags in the VAR. Simply put, they do not work in a reasonable amount of time. More precisely, when we run the sampler for a large number of replications, only a handful of draws are accepted. For instance, in a long run of post-burn in replications, 99.97% of the candidate draws had at least one state which violated the restriction and, hence, fewer than 0.03% of candidate draws were accepted in the Metropolis-Hastings algorithm. This shows how, in serious macroeconomic applications, it is very easy for the VAR coefficients to evolve in ways that allow for some posterior probability to lie in explosive regions. Insofar as one believes such explosive behavior is unrealistic, it is desirable to rule it out in the prior. This further motivates why the issues discussed in this paper are important ones. In a case like this, where multi-move algorithms reject virtually every draw, the researcher would have no choice but to use our single move algorithm.

In the remainder of this paper we use one lag in our VAR. This more parsimonious choice yields more precise inference and less evidence of explosive behavior in the VAR coefficients. We begin with a discussion of the computational performance of the various algorithms. The key issue with these algorithms is the speed of mixing. To evaluate this issue, we present the commonly-used effective sample size (see, among many others, Holmes and Held, 2006), defined as:

16

$$ESS = \frac{S}{1 + 2\sum_{h=1}^{H} \rho(h)}$$

where S is the number of (post burn-in) MCMC replications, $\rho(h)$ is the autocorrelation between MCMC draws h replications apart. ESS can be calculated for any feature of interest. Given that the key issues in this paper relate to the time varying VAR coefficients we base our features of interest on these. To keep things manageable we use $\sum_{t=1}^{T} \theta_t$ as our feature of interest and present ESS for one element of this vector. In Table 1 we choose the coefficient on the lagged interest rate in the interest rate equation (but results for other coefficients are similar). Note that, if the MCMC draws were independent of one another, we would get $\sum_{h=1}^{H} \rho(h) \approx 0$ and $ESS = S$. With positively correlated MCMC draws we get $ESS < S$ and the degree to which ESS is less than S is a measure of how slowly mixing the algorithm is. Table 1 is produced with $S = 50,000$ (following $10,000$ burn-in draws) using $H = 100$. We use our single and multi-move algorithms as well as the variants of both algorithms with $R(\theta_{t-1}, Q) \approx c$ (which we call the "Cogley-Sargent Approximation"). Table 1 also presents the proportion of candidate draws that are accepted (in the column labelled "Prop. Switch") in the algorithm for drawing the states.

The top third of Table 1 presents some evidence on standard algorithms without imposing inequality restrictions on states. Note that these algorithms simply use the unrestricted posteriors. It can be seen that, even without the complications caused by the imposition of inequality restrictions, these algorithms can be very slow to mix. As we would expect, the multi-move algorithm has an effective sample size which is almost twice as high as the single move algorithm. But it still mixes very slowly. A rough rule of thumb would say that, to achieve the same degree of accuracy as $S = 10,000$ independent draws, the multi-move algorithm would require $S = \frac{10,000}{0.0283} > 300,000$ draws. This finding serves as a warning to researchers using Bayesian methods for TVP-VARs with multivariate stochastic volatility. Care should be taken in monitoring convergence of MCMC algorithms.

Once we impose restrictions on the states, it is likely that the mixing of the MCMC algorithms will get worse due to the Metropolis-Hastings algorithm not accepting all candidate draws. Although there is a contradictory tendency in that allowing for explosive VAR coefficients can slow mixing so that, by ruling these out mixing can improve. The questions we want to

address are: "how much worse/better do the algorithms become when restrictions are imposed?" and "by how much does the performance of single move algorithms improve relative to multi-move variants when we impose restrictions on the states?". An examination of Table 1 tells a strong and consistent story: With the multi-move algorithm, imposing inequality restrictions causes a large decrease in efficiency. However, for the single-move algorithms, adding inequality restrictions does not cause the (admittedly initially poor) mixing performance of the unrestricted posterior to decline.

With the Primiceri prior, the imposition of the inequality restrictions leads to multi-move algorithms where about 10% of the Metropolis-Hastings algorithm draws are accepted. This causes the effective sample size to fall substantially. With the Alternative Prior, which allows for more variation in the VAR coefficients, the multi-move algorithm accepts almost no post burn-in candidate draws. Hence, virtually all the draws of θ_t are exactly the same as one another and, thus, are almost perfectly correlated with one another. Thus, to all intents and purposes, the multi-move algorithm is computationally infeasible with the Alternative Prior.

The single move algorithm, in contrast, always works roughly the same in all cases. For this algorithm there are T Metropolis-Hastings candidate generating densities and, hence, "Prop. Switch" is the average of these. It can be seen that the single-move algorithm does continue to accept candidate draws and, thus, its performance does not deteriorate when inequality restrictions are imposed.

Another point worth noting is that (for a given number of replications) our Metropolis-Hastings algorithms work slightly better than the Cogley-Sargent approximations. This is partly due to our way of drawing Q. That is, the Cogley-Sargent approximation and our algorithm both take candidate draws of Q from the unrestricted posterior. However, the Cogley-Sargent approximation accepts every draw, where we only accept them with probability α_Q given in (9). This helps keep the candidate draws of θ^T in the non-explosive region.

For completeness, we note that, when using Primiceri's prior, the proportion of accepted draws of Q is 0.4673 (for the single move algorithm) and 0.4641 (for the multi-move algorithm). For the Alternative Prior these numbers are 0.4953 and 0.6670, respectively.

Table 1: Evidence on Mixing of Algorithms				
Algorithm	No Restrictions on the States			
	Primiceri Prior		Alternative Prior	
	$\frac{ESS}{S}$	Prop. Switch	$\frac{ESS}{S}$	Prop. Switch
Single move	0.0092	1.0000	0.0063	1.0000
Multi-move	0.0283	1.0000	0.0138	1.0000
	Restrictions on States			
	Primiceri Prior		Alternative Prior	
	$\frac{ESS}{S}$	Prop. Switch	$\frac{ESS}{S}$	Prop. Switch
Single move	0.0168	0.7981	0.0143	0.8387
Multi-move	0.0098	0.1362	0.0058	0.0047
	Primiceri Prior		Alternative Prior	
	$\frac{ESS}{S}$	Prop. Switch	$\frac{ESS}{S}$	Prop. Switch
Single move Cogley-Sargent Approx.	0.0103	0.8351	0.0073	0.7877
Multi-move Cogley-Sargent Approx.	0.0062	0.1092	0.0050	0.0007

As discussed in Section 2, our algorithms require the evaluation of key integrating constants which we have labelled $R(\theta_{t-1}, Q)$ for the multi-move algorithms. For the present case, where we are using a complicated, nonlinear inequality restriction, evaluation of these integrating constants requires the use of simulation methods which can slow down the algorithm. As we have seen, this has led other authors such as Cogley and Sargent (2005) to ignore these integrating constants, implicitly assuming that $R(\theta_{t-1}, Q)$ is approximately constant. If it is not approximately constant then the approximation inherent in their MCMC algorithm could be a poor one. In addition, remember that $\prod_{t=1}^{T} R(\theta_{t-1}, Q)$ appears in (10), the formula for Bayes factor comparing the restricted to the unrestricted model. Thus, inappropriate treatment of $\prod_{t=1}^{T} R(\theta_{t-1}, Q)$ can potentially lead to misleading inference about how probable the restrictions are.

Of course, the question of whether the $R(\theta_{t-1}, Q) \approx c$ approximation is a good one is application-specific. Nevertheless, we feel our TVP-VAR is representative of the sorts of applications used in practice and should shed some light on whether the algorithms proposed in this paper are necessary, as opposed to the simpler approximate algorithm of Cogley and Sargent (2005). Figure 1 plots a histogram of the acceptance probabilities, α_θ, which depend on $\prod_{t=1}^{T} R(\theta_{t-1}, Q)$ evaluated at old and candidate draws. It is based on all

the (post burn-in) MCMC draws which satisfy the inequality restrictions (the remaining draws have acceptance probability of zero). It uses our multi-move algorithm and the Primiceri prior.

Figure 1 shows how there is a large degree of variation in this acceptance probability across MCMC draws. The Cogley-Sargent approximate algorithm would simply accept all these draws.[6] Clearly our algorithm has acceptance probabilities which can be quite different from the alternative of either rejecting or accepting every candidate draw depending only on whether any θ_t exhibits explosive behavior or not. These facts suggest that assuming $\prod_{t=1}^{T} R(\theta_{t-1}, Q)$ to be constant across draws (as approximate methods do) can be misleading.

So far, we have compared algorithms in terms of their mixing properties and Metropolis-Hastings acceptance probabilities. Do the algorithms differ enough to have substantial implications for empirical results? This depends partly on how empirical results are presented. For instance, if there is a small amount of posterior probability in the explosive region for θ_t, posterior means and standard deviations will be much more effected by this than posterior medians and inter-quartile ranges. To give an example of the practical implications of the issues addressed in this paper, consider the Bayes factor comparing the restricted model to the unrestricted model. This can be calculated using the MCMC draws from the unrestricted posterior as described in (11). Remember that this Bayes factor involves two terms, one involving the integrating constant $\prod_{t=1}^{T} R(\theta_{t-1}, Q)$ and the other involving the proportion of draws which satisfy the restrictions (the latter we called Pr [Restrictions are true]). In this data set, using the Primiceri prior, we find Pr [Restrictions are true] $= 0.0088$.[7] Some might use this latter as very strong evidence against the restrictions interpreting it (incorrectly) as saying there is less than 1% probability that the restrictions hold. However, when we correctly calculate the Bayes factor using (11), we find it to be 0.1977. This still suggests that the restrictions are unlikely to be true, but to a far lesser extent. That is, the unrestricted model is only five times as proba-

---

[6]This statement is only approximately correct since the Cogley-Sargent algorithm differs from ours in its treatment of Q.

[7]Remember that this number is calculated using the unrestricted posterior, not the candidate generating density in our multi-move algorithms. Since the unrestricted posterior draws Q conditional on all draws of θ^T, whereas our multi-move algorithms draw Q conditional on accepted draws of θ^T, this number is not the same as the proportion of switches reported in Table 1.

Figure 1: Histogram of M-H Acceptance Probs. for θ
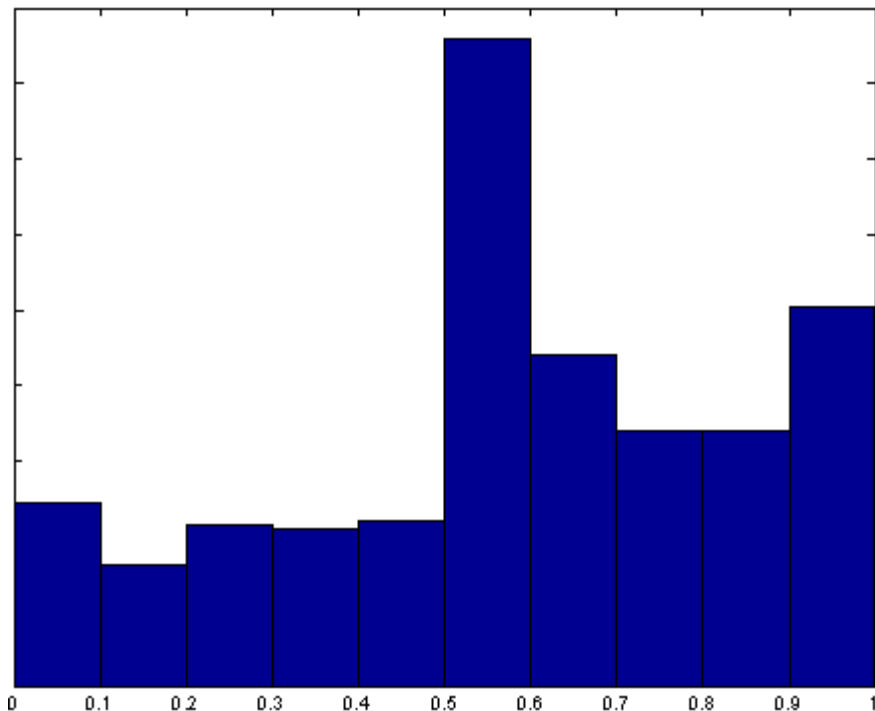
ble as the restricted one. Alternatively, using the popular Kass and Raftery (1995) rule of thumb for interpreting Bayes factors, the Bayes factor lies in the region where the recommended conclusion is "slight evidence against the restrictions". Clearly, proper treatment of the integrating constant can be of substantial empirical importance in real world applications.

4   Conclusions

In this paper, we have developed algorithms for imposing inequality restrictions on the states in state space models. Our methods are quite general, but we have focussed on using them in TVP-VAR models. In particular, we consider the important empirical issue of imposing the restriction that the roots of the matrix polynomial defined by the VAR coefficients at each point in time all lie outside the unit circle. There are two problems with conventional algorithms for imposing such restrictions. Firstly, they are multi-move algo-

rithms which involve simply discarding candidate draws of the entire state vector when the VAR coefficients at any point in time violate the inequality constraint. In this paper, we have shown how it can easily arise in practice that virtually all the candidate draws end up being discarded and, to all intents and purposes, such conventional algorithms simply do not work given finite computing power. To get around this problem, we derive a single move algorithm which discards only the candidate draw at time t if the inequality constraint is violated. Typically, researchers avoid single move algorithms for state space models since they mix more slowly than multi-move algorithms. However, in our empirical work, we find that the cost imposed by slower mixing is relatively small. Furthermore, our single move algorithm works in cases where the multi-move algorithm fails.

A second problem with conventional algorithms is that they are approximate, ignoring a certain integrating constant. This integrating constant can be awkward to calculate (since it often requires the use of simulation methods). But nevertheless, our empirical investigations indicate that it is important to include it since it can vary substantially across MCMC draws and have a substantive impact of results.

References

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

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.

Cooley, T. and Prescott, E., 1973, Varying parameter regressions: A theory and some applications, *Annals of Economic and Social Measurement*, 2, 463-473.

Cooley, T. and Prescott, E., 1973, Estimation in the presence of stochastic parameter variation, *Econometrica*, 167-184.

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

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

Geweke, J., 1991, Efficient simulation from the multivariate Normal and Student-t distributions subject to linear constraints, pages 571-578 in E. Keramidas (ed.), *Computer Science and Statistics: Proceedings of the Twenty-Third Symposium on the Interface.* Fairfax: Interface Foundation of North America, Inc..

Holmes, C. and Held, L., 2006, Bayesian auxiliary variable models for binary and multinomial regression, *Bayesian Analysis*, 1, 145-168.

Kim, C.-J. and Nelson, C., 1999, *State-Space Models with Regime Switching: Classical and Gibbs Sampling Approaches with Applications*, 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.

Kass, R. and Raftery, A., 1995, Bayes factors, *Journal of the American Statistical Association*, 90, 773-795.

Koop, G. , 2003, *Bayesian Econometrics,* Chichester: Wiley.

Koop, G., Leon-Gonzalez, R. and Strachan, R., 2007, On the evolution of monetary policy, manuscript available at http://personal.strath.ac.uk/gary.koop/.

Koop, G., Leon-Gonzalez, R. and Strachan, R., 2008, Dynamic probabilities of restrictions in state space models: An application to the Phillips curve, manuscript available at http://personal.strath.ac.uk/gary.koop/.

Koop, G. and Potter, S., 2008, A new model for underlying inflation, manuscript available at http://personal.strath.ac.uk/gary.koop/.

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

Stock, J. and Watson, M., 2007, Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39, 3-33.

Appendix: Treatment of Measurement Error Covariance Matrix and Additional Details about Priors

In the body of the paper, we left H_t unspecified. This appendix provides details about how this is treated in our empirical work. The approach is the same as in Primiceri (2005) so we offer only the general outline and refer the reader to Primiceri's paper for exact formulae for the relevant blocks of the MCMC algorithm. We begin with a triangular reduction of H_t such that:

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

or

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

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

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

For Σ_t a stochastic volatility framework is used. In particular, if $\sigma_t = (\sigma_{1,t}, .., \sigma_{p,t})'$, $h_{i,t} = \ln(\sigma_{i,t})$, $h_t = (h_{1,t}, .., h_{p,t})'$ then we use:

$$h_{t+1} = h_t + u_t,$$

where u_t is $N(0, W)$ and is independent over t and of ε_t and η_t. The algorithm of Kim, Shephard and Chib (1998) can be used to draw these volatilities.

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

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

where ζ_t is $N(0, C)$ and is independent over t and of u_t, ε_t and η_t. . Following Primiceri (2005), we assume C to have a block diagonal structure such that the coefficients in C belonging to each equation are independent of one another. With regards to our MCMC algorithm, this means that we can transform the original measurement equation so that the Durbin and

Koopman (2002) algorithm can be used to draw the states one equation at a time.

Posterior draws W and the blocks of C involve the standard Wishart forms. Full details are given in Primiceri (2005) or Koop, Leon-Gonzalez and Strachan (2007). With respect to the priors for W and C as well as the initial conditions in the two state equations that appear in the volatility specification, suffice it to note here that we use exactly the same training sample prior as in Primiceri (2005) or Koop, Leon-Gonzalez and Strachan (2007).

We treat θ_0 as a regression effect and, following Primiceri (2005), use a prior of the form:

$$\theta_0 \sim N\left(\widehat{\theta}, 4\widehat{V}\right),$$

where $\widehat{V}$ is defined in the body of the paper.