STRATHCLYDE

DISCUSSION PAPERS IN ECONOMICS

MODELLING BREAKS AND CLUSTERS IN THE STEADY STATES OF MACROECONOMIC VARIABLES

BY

JOSHUA C.C. CHAN AND GARY KOOP

NO. 11-11

DEPARTMENT OF ECONOMICS
UNIVERSITY OF STRATHCLYDE
GLASGOW

Modelling Breaks and Clusters in the Steady States of Macroeconomic Variables

Joshua C.C. Chan, Australian National University*
Gary Koop, University of Strathclyde†

January 2011

Abstract

Macroeconomists working with multivariate models typically face uncertainty over which (if any) of their variables have long run steady states which are subject to breaks. Furthermore, the nature of the break process is often unknown. In this paper, we draw on methods from the Bayesian clustering literature to develop an econometric methodology which: i) finds groups of variables which have the same number of breaks; and ii) determines the nature of the break process within each group. We present an application involving a five-variate steady-state VAR.

1 Introduction

Macroeconomists working with multivariate models such as VARs face a myriad of modelling choices. Traditionally, such choices have involved restrictions on parameters. For instance, cointegration, lag length selection or the economic theory used by DSGE modelers all involve restrictions on the coefficients of a VAR (or similar multivariate time series model). However, the increasing realization of the importance of parameter change has led

*Joshua Chan would like to acknowledge financial support from the Australian Research Council under Grant DP0987170.

†Gary Koop thanks the ESRC for support under grant RES-062-23-2646. He is a Fellow at the Rimini Centre for Economic Analysis.

macroeconomists to work with more parameter-rich models which allow for such change. Examples include time-varying parameter (TVP) VAR models (see, among many others, Cogley and Sargent, 2001, 2005 or Primiceri, 2005), multivariate Markov switching models such as Sims and Zha (2006) or structural break VAR models such as Jochmann, Koop and Strachan (2010).

Often the researcher is unsure of the nature of parameter change (e.g. is it associated with VAR coefficients or the error covariance matrix? is it associated with time such as in a structural break model or does change occur over the business cycle?, etc.). Multivariate time series models such as VARs are parameter-rich even with constant parameters. Allowing for parameter change in VARs increases the number of parameters to be estimated. This raises worries about over-fitting and over-parameterization. The presence of model uncertainty relating to time-variation in parameters greatly exacerbates these worries.

The present paper is motivated by these considerations. Faced with uncertainty over the nature of parameter change, we want an econometric method which will discover its nature in a data-based fashion. And faced with overfitting, we want to do this in as parsimonious manner as possible. In many cases, this latter goal can be achieved by focussing on economically important parameters. For instance, a VAR may have hundreds of parameters (or even more, see Banbura, Giannone and Reichlin, 2010). These parameters control the dynamics (short-run and long-run) of the variables in the model as well as the economic relationships of interest to the macroeconomists (e.g. impulse responses are functions of VAR coefficients and the error covariance matrix). But VAR coefficients are hard to directly interpret and allowing for parameter change in all of them can lead to a very parameter-rich model. When considering ways of allowing for parameter change, the researcher may wish to focus on some economically meaningful function of the parameters (e.g. allowing for only the monetary transmission mechanism to change). And it is typically most important to model parameter change in the economic feature under study. For instance, in a study of the monetary policy transmission mechanism it is very important to correctly model parameter change in this relationship, but it may be less important to correctly model parameter change in other parts of the model.

In this paper, we develop an econometric methodology which is more parsimonious than other approaches (such as TVP-VARs) and uncovers parameter change of an unknown sort in features of economic importance. We focus on the long run steady states of VAR dependent variables (although the

general ideas can be adapted to any feature of interest). These are features that have a straightforward economic interpretation and theoretical macroeconomic models such as DSGE models typically have strong implications for long run steady states. We extend the steady state VAR of Villani (2009) to allow for the steady states to change over time. Of course, it would be straightforward to adapt any of the existing modelling approaches described above (e.g. Markov switching or structural break models) so as to apply only to the steady states. However, such an approach would assume all of the steady states change in a particular way (e.g. a structural break model would imply they all change at the break time). The econometric methodology developed in this paper (drawing on ideas from the Bayesian clustering literature, see Tadesse, Sha and Vannucci, 2005) is more sophisticated than this. It determines (in an automatic, data-based fashion) which variables exhibit breaks in their steady states (i.e. some variables can exhibit breaks and others not) and the nature of the break process (e.g. it can estimate structural breaks which occur at a point in time or parameter change over the business cycle or anything else).

The paper is organized as follows. The next section of this paper describes the modelling framework and provides a general outline of the Markov chain Monte Carlo (MCMC) algorithm used to estimate the model. Full technical details on prior, posterior and MCMC algorithm are provided in the Technical Appendix. The third section of the paper illustrates the usefulness of our methods in an empirical application relating to one presented in Del Negro and Schorfheide (2008). We use a five-variate VAR and find that breaks exist in the steady states of some of the series but not others.

2 Modelling Framework

Bayesian VAR analysis traditionally works with a VAR of the form:

$$A^{\dagger}(L)\, y_t = \mu^{\dagger} + \varepsilon_t$$

where y_t is an $n \times 1$ vector of dependent variables for $t = 1, ..., T$, ε_t is $N(0, \Sigma)$ and $A^{\dagger}(L) = I - A_1^{\dagger} L - .. - A_p^{\dagger} L^p$ is a polynomial in the lag operator. Conventional Bayesian VAR approaches such as the Minnesota prior (see Doan, Litterman and Sims, 1984) place a prior on the parameters in $A^{\dagger}(L)$ and $\mu^{\dagger}$. This parameterization can be hard to directly interpret (e.g. $\mu^{\dagger}$ is

not the unconditional mean of the series). In contrast to this, the steady state VAR (see Villani, 2009) can be written as

$$A(L)(y_t - \mu) = \varepsilon_t. \tag{1}$$

This specification for the VAR has the advantage that μ is the unconditional mean of y_t and, thus, can be interpreted as the steady state of y_t. As argued in Villani (2009), the steady state is often something that researchers have strong prior beliefs about (unlike $A_1, .., A_p$). Thus, it may be preferable to focus prior elicitation efforts on μ. The parameters in $A(L)$, controlling the short run dynamics for deviations from steady states, may be of less interest to the macroeconomist. For instance, DSGE modelers often have strong prior information about steady states and elicit their priors in terms of such structural parameters (see, among many others, Smets and Wouters, 2007 and Del Negro and Schorfheide, 2008). A drawback of the steady state VAR relative to the traditional VAR is that MCMC methods must be used. However, the gain in interpretability and the ability to elicit priors directly off of parameters with an economic interpretation are large benefits which may outweigh this drawback.

In empirical macroeconomic work, it is likely that the steady states of some variables remain constant over time, while others change at a particular point in time (i.e. structural breaks might occur), while others might change in some other fashion (e.g. they may differ between expansions and recessions). However, the researcher is typically unsure about which of these possibilities holds for which variable. Unless n is small, the number of modelling choices can be daunting. In this paper, we draw on ideas from the Bayesian clustering literature (see, e.g., Tadesse, Sha and Vannucci, 2005) to propose a modelling framework which allows us to group the dependent variables into clusters which have the same structure. For instance, one cluster might have constant steady states, another cluster might include dependent variables whose long run steady states exhibit a break, etc. This grouping is done in an automatic data-based manner.

Since the contributions of this paper relate to μ, we will draw out the basic intuition of our methodology ignoring the role of $A(L)$ and the error covariance matrix. Of course, in our empirical application $A(L)$ will be included as well as a time-varying error covariance matrix. Complete details of the full model are given in the Technical Appendix.

Accordingly, let us begin with a simple model:

$$y_t = \mu_1 + \varepsilon_t \tag{2}$$

where ε_t is $N(0, \Sigma_1)$. In the spirit of Tadesse, Sha and Vannucci (2005), we begin by extending this to a mixture of Normals specification where

$$p(y_t|q, \mu, \Sigma) = \sum_{j=1}^{G} q_j \phi\left(y_t|\mu_j, \Sigma_j\right), \tag{3}$$

where $\phi\left(y_t|\mu_j, \Sigma_j\right)$ is the p.d.f. of the Normal distribution with mean μ_j and variance Σ_j, $q = (q_1, .., q_G)'$, $\mu = (\mu_1', .., \mu_G')'$ and $\Sigma = (\Sigma_1, .., \Sigma_G)$. Or, equivalently, we can introduce the discrete random variables: $\lambda_t \in \{1, .., G\}$ such that:

$$y_t|\lambda_t = j \sim N\left(\mu_j, \Sigma_j\right) \tag{4}$$

where $p(\lambda_t = j) = q_j$. For future reference, let $\lambda = (\lambda_1, .., \lambda_T)'$ and $y = (y_1', .., y_T')'$.

This is a standard mixture of Normals representation which has been used in many papers. Mixtures of Normals are very flexible as discussed, e.g., in Geweke and Keane (2007) and Geweke and Amisano (2011). For our purposes, we stress that (4) allows for clustering over time and so would be able to pick up features like structural breaks or regime switching, where the steady states of all variables change. That is, it says each y_t for $t = 1, .., T$ can be drawn from one of G different distributions. For instance, if $y_1, .., y_\tau$ were drawn from one distribution and $y_{\tau+1}, .., y_T$ were drawn from a second, then we would have a structural break at time τ. But it is also possible that y_j is drawn from the first distribution where j denotes times when the economy is in recession (and other time periods are drawn from a second distribution). Then we have a model with properties similar to a Markov switching model where properties differ over the business cycle. In general, any grouping is possible.

It is also worth noting that (4) allows for the error covariance matrix to differ across regimes. So formally, if we find evidence for $G > 1$ this implies either that the steady states are changing or that the error covariance matrix is changing. It is possible to restrict $\Sigma_1 = .. = \Sigma_G$ if the researcher wishes to focus solely on steady-state changes in the context of a homoskedastic VAR. Alternatively, at the cost of adding extra blocks to the MCMC algorithm, the error covariance matrix could be modelled separately from the mixtures of

Normal component of the model (e.g. as a multivariate stochastic volatility process).

Simple mixtures of Normals model such as (4) allow for all parameters to differ across elements in the mixture. For high-dimensional models such as VARs such flexibility can lead to an over-parameterized model. This flexibility can be unnecessary and lead to undesirable consequences. If the steady states in only one or two variables change and n is large, it is distinctly possible that the econometric model will indicate no change. That is, the model (4) offers the choice between "steady states of all variables are constant over time" and "all n steady states change". Given these alternatives, and the reward for parsimony built into Bayesian model selection methods, it will only choose the latter if most of the steady states change or if the change in the steady state in one variable is huge. These considerations motivate the development of a model designed to pick up breaks which occur only in some subset of the variables. In practice, the researcher rarely knows which subsets of variables might have breaks in their steady states, so the model should be able to find these subsets in an automatic data-based fashion. The following extension of the standard mixture of Normals model can achieve these goals.

Let $\gamma = (\gamma_1, .., \gamma_n)'$ be a vector of dummy variables where $\gamma_j = 1$ implies that the j^{th} dependent variable follows a mixture of Normals representation such as that given in (4). If $\gamma_j = 0$ then the j^{th} dependent variable does not follow a mixture of Normals, but rather has a time-invariant steady state (and error covariance matrix) such as (2). In other words, γ serves to divide our dependent variables into two groups, where one group allows for up to G changes in steady states over time or across regimes and the other group has constant steady states.

Formally, let $y_{t(\gamma=1)}$ denote the vector containing the elements of y_t which have $\gamma_j = 1$ and $y_{t(\gamma=0)}$ the vector containing the remaining elements of y_t. And adopt the same "subscript $(\gamma = 0/1)$" notational convention for the parameters in the model (e.g. $\mu_{(\gamma=0)}$ will be the unconditional mean for all variables which have a time-invariant steady state, $n_{(\gamma=0)}$ will be the number of such variables, etc.). This leads to the following distribution for y, λ which Tadesse et al (2005) use as their likelihood function:[1]

[1]In the terminology of, e.g., Fruhwirth-Schnatter and Wagner (2008) this is the complete data likelihood, as opposed to the integrated likelihood which would be $p(y|\gamma, q, \mu, \Sigma)$.

$$
\begin{aligned}
&p\left(y,\lambda|\gamma,q,\mu,\Sigma\right)=2\pi^{-n_{(\gamma=0)}\frac{T}{2}}\left|\Sigma_{(\gamma=0)}\right|^{-\frac{T}{2}}\\
&\times\exp\left\{-\tfrac{1}{2}\sum_{t=1}^{T}\left(y_{t(\gamma=0)}-\mu_{(\gamma=0)}\right)'\Sigma_{(\gamma=0)}^{-1}\left(y_{t(\gamma=0)}-\mu_{(\gamma=0)}\right)\right\}\\
&\times\prod_{j=1}^{G}2\pi^{-n_{(\gamma=1)}\frac{T_j}{2}}\left|\Sigma_{j,(\gamma=1)}\right|^{-\frac{T_j}{2}}q_j^{T_j}\\
&\times\exp\left\{-\tfrac{1}{2}\sum_{t\in C_j}\left(y_{t(\gamma=1)}-\mu_{j,(\gamma=1)}\right)'\Sigma_{(j,\gamma=1)}^{-1}\left(y_{t(\gamma=1)}-\mu_{j,(\gamma=1)}\right)\right\}
\end{aligned}
\tag{5}
$$

where T_j is the number of observations in the j^{th} element in the Normal mixture given in (4) and $t\in C_j$ denotes observations belonging to cluster j (i.e. C_j is the set of observations for which $\lambda_t=j$). Note that λ enters this likelihood function since it determines T_j.

To carry out posterior simulation in the steady state VAR version of this model, we require a prior to combine with the likelihood function (5) and an MCMC algorithm for drawing the parameters $\gamma,q,\mu,\Sigma,\lambda$ and $A_1,..,A_p$. Note that the relationship $p\left(\lambda_t=j\right)=q_j$ given after (4) provides us with a hierarchical prior for λ. For γ we assume a Bernoulli prior which implies, a priori, each variable is equally likely to exhibit no breaks as exhibit breaks. Complete details of these priors and priors for other parameters are given in the Technical Appendix.

The basic idea of our MCMC algorithm can be described very simply: it combines the algorithm of Tadesse et al (2005) with a Bayesian VAR algorithm. Complete details are given in the Technical Appendix. Finally, we choose G using the Bayesian information criterion (BIC) as detailed in the Technical Appendix.

3 Empirical Illustration with a Five-variate VAR

3.1 Data

The data set is obtained from Del Negro and Schorfheide (2008). It runs from 1954Q3 through 2005Q4 and consists of five commonly-used US macroeconomic variables: GDP per capita (Y_t), hours worked per capita (L_t), labor

income share (lsh_t), the GDP deflator (P_t) and interest rates (the Fed Funds rate, R_t). The variables are constructed and transformed exactly as in Del Negro and Schorheide (2008). The definitions of the variables which enter the VAR are reproduced below:

Output growth	$400(\ln Y_t - \ln Y_{t-1})$
Hours	$100 \ln L_t$
Labor share	$100 \ln lsh_t$
Inflation	$400(\ln P_t - \ln P_{t-1})$
Interest rates	$400 \ln R_t$

3.2 Empirical Illustration

Table 1 presents BICs for various values of G and p and it can be seen that $G = 2$ and $p = 2$ are the preferred choices. There is strong support for models with breaks in steady states (since $G = 1$ receives little support), but a small number of breaks seems adequate (since there is little evidence for $G = 3$). In the remainder of this section, we set $G = p = 2$.

Table 1: BIC's.			
	$G = 1$	$G = 2$	$G = 3$
$p = 1$	2814.2	2823.7	2872.5
$p = 2$	2755.3	2742.2	2798.3
$p = 3$	2885.3	2875.0	2931.2

The evidence in favor of $G = 2$ suggests breaks are occurring in some of the steady states. But which ones? The estimated posterior mode for the cluster label, γ, is $(0, 0, 0, 1, 1)'$, suggesting only inflation and interest rates experienced changes in their steady states. The posterior mean is very similar to the posterior mode, indicating the clustering algorithm is clearly identifying which variables have breaks and which ones do not. If we re-run the MCMC algorithm, conditional on $\gamma = (0, 0, 0, 1, 1)'$, we can obtain parameter estimates which arise from a single model with a clear interpretation. This model is parameterized in terms of $\mu_{(\gamma=0)} = (\mu^1_{(\gamma=0)}, \mu^2_{(\gamma=0)}, \mu^3_{(\gamma=0)})'$ (i.e. a 3×1 vector with elements being the steady states for output growth per capita, hours worked per capita and labor income share respectively), $\left(\mu^1_{1,(\gamma=1)}, \mu^1_{2,(\gamma=1)}\right)$ (i.e. these are the steady states in the two regimes for inflation) and $\left(\mu^2_{1,(\gamma=1)}, \mu^2_{2,(\gamma=1)}\right)$ (i.e. these are the steady states in the two

regimes for interest rates). The error covariances are labelled with a similar notational convention where subscripts denote regimes and superscripts particular parameters within a regime. The estimated posterior means and standard deviations for some of the key parameters are reported in Table 2.

Most of the parameter estimates are similar to those found in a standard VAR, so we will focus our discussion on those which differ. Table 2 provides strong evidence of a large change in the steady state for inflation, with slightly weaker evidence in favor of a smaller change in the steady state for the interest rate. But note also that the regime change we are finding looks to be associated with a change in the volatility of both variables. Remember that our approach allows for different regimes to have different means and different error variances. Clearly, this latter aspect is important in this data set. For both inflation and interest rates we are finding volatilities to be much lower in the second regime than the first regime. Since the point estimate of $q_1 = 0.071$ the high volatility regime occurs much less frequently than the low volatility regime.

Table 2: Posterior Means and Standard Deviations of Parameters

parameter	$\mathbb{E}(\cdot\|y)$	$\sqrt{\text{Var}(\cdot\|y)}$	parameter	$\mathbb{E}(\cdot\|y)$	$\sqrt{\text{Var}(\cdot\|y)}$
$\mu^1_{(\gamma=0)}$	0.449	0.042	q_1	0.071	0.030
$\mu^2_{(\gamma=0)}$	552.6	8.954	$\Sigma^{1,1}_{(\gamma=0)}$	0.827	0.085
$\mu^3_{(\gamma=0)}$	−55.44	0.911	$\Sigma^{2,1}_{(\gamma=0)}$	0.466	0.059
$\mu^1_{1,(\gamma=1)}$	6.969	1.176	$\Sigma^{3,1}_{(\gamma=0)}$	−0.967	0.051
$\mu^2_{1,(\gamma=1)}$	6.938	1.883	$\Sigma^{2,2}_{(\gamma=0)}$	0.525	0.055
$\mu^1_{2,(\gamma=1)}$	3.377	0.240	$\Sigma^{3,2}_{(\gamma=0)}$	−0.101	0.037
$\mu^2_{2,(\gamma=1)}$	5.658	0.467	$\Sigma^{3,3}_{(\gamma=0)}$	0.452	0.046
$\Sigma^{1,1}_{1,(\gamma=1)}$	5.157	2.378	$\Sigma^{1,1}_{2,(\gamma=1)}$	2.288	0.259
$\Sigma^{2,1}_{1,(\gamma=1)}$	0.323	1.730	$\Sigma^{2,1}_{2,(\gamma=1)}$	−0.162	0.142
$\Sigma^{2,2}_{1,(\gamma=1)}$	6.975	3.168	$\Sigma^{2,2}_{2,(\gamma=1)}$	0.689	0.110

Figure 1 presents evidence on when the two regimes are occurring. That is, it reports the estimated probabilities $P(\lambda_t = 1|y)$ for $t = 1, \ldots, T$. Of course, the reader can figure out $P(\lambda_t = 2|y)$ since it is $1 - P(\lambda_t = 1|y)$. It can be seen that our econometric methodology is clearly finding changes in steady states and volatilities in interest rates and inflation to be associated with the period 1973-1983 (i.e. outside this time there is strong evidence that

the second regime holds). This is the period after the collapse of the Bretton-Woods agreement and OPEC oil price shock up until the great moderation of the business cycle. Of course, macroeconomists could debate as to whether the changes in steady states we are finding truly were long run changes (as opposed to persistent responses to exogenous shocks). Nevertheless, the pattern we have flagged in Figure 1 is a sensible one. And it is interesting to note that it does not apply to the real variables in our model, but the nominal variables: the interest rate and inflation.

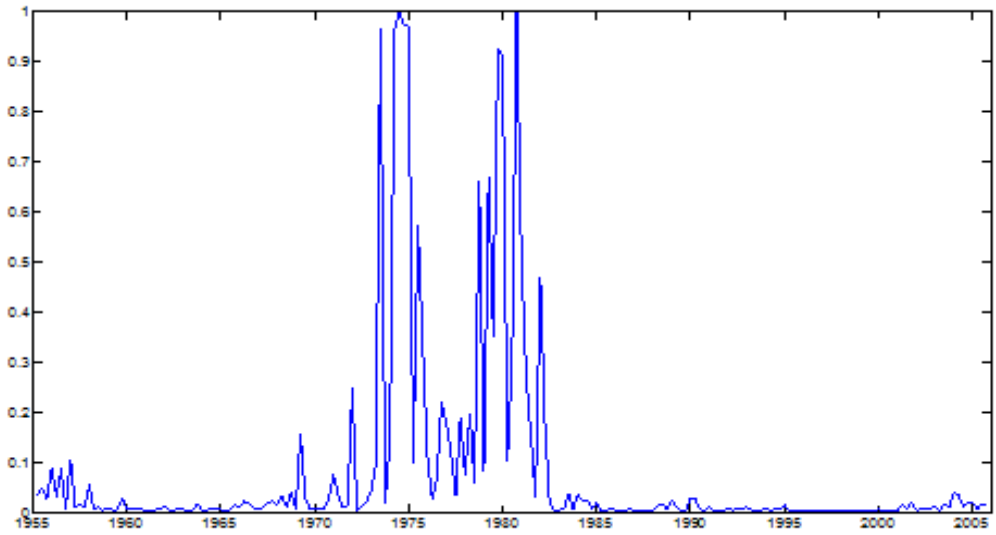

Figure 1: Probability of First Regime: $\mathbb{P}(\lambda_t = 1 \,|\, y)$.

4 Conclusion

In this paper, we have developed an econometric methodology for multivariate macroeconomic models, based on Tadesse et al (2005), which differs from existing methods in that it allows both for clustering in terms of variables and in terms of regimes. That is, it automatically divides the variables into groups. Within each group, variables exhibit a common pattern (e.g. they

can exhibit structural breaks at the same time). We apply this methodology to the case of the steady-state VAR of Villani (2009). We focus on the issue of breaks in long-run steady states, although we argue that the methodology could be useful in a wide variety of empirical macroeconomic contexts.

Our empirical illustration, using a moderately-sized VAR, indicates that the methodology works well and leads to parsimonious representations. Instead of allowing for breaks in the steady states (and error covariance matrices) of all variables (as would be done in a conventional structural break model) or allowing for breaks in dozens or hundreds of VAR coefficients (as would be done in a Markov-switching VAR or TVP-VAR), our methodology indicates breaks are occurring in the steady states of only two variables.

References

Banbura, M., Giannone, D. and Reichlin, L. (2010). "Large Bayesian VARs," *Journal of Applied Econometrics*, 25, 71-92.

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.

Del Negro, M. and F. Schorfheide (2008). "Forming priors for DSGE models (and how it affects the assessment of nominal rigidities)," *Journal of Monetary Economics*, 55, 1191-1208.

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

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.

Geweke, J. and Amisano, G. (2011). "Hierarchical Markov normal mixture models with applications to financial asset returns," *Journal of Applied Econometrics*, 26, 1-29.

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

Jochmann, M., Koop, G. and Strachan, R. (2010). "Bayesian forecasting using stochastic search variable Selection in a VAR Subject to Breaks," International Journal of Forecasting, 2010, 26, 326-347.

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

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

Smets, F. and Wouters, R. (2007). "Shocks and frictions in US business cycles: A Bayesian DSGE approach," *American Economic Review*, 97, 588-606.

Tadesse, M., Sha, N. and Vannucci, M. (2005). "Bayesian variable selection in clustering high-dimensional data," *Journal of the American Statistical Association*, 100, 602-617.

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

Technical Appendix

The Model

Write the steady state VAR as:

$$(y_t - \mu_{\lambda_t}) - A_1(y_{t-1} - \mu_{\lambda_t}) - \cdots - A_p(y_{t-p} - \mu_{\lambda_t}) = \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_{\lambda_t}), \tag{6}$$

where $\lambda_t \in \{1, 2, \ldots, G\}$ denotes regimes and $A_i, i = 1 \ldots, p$ are $n \times n$ matrices of VAR coefficients such that y_t is stationary (the companion matrix of $A = (A_1, \ldots, A_p)$ has roots outside the unit circle). Let $\gamma = (\gamma_1, \ldots, \gamma_n)'$ be a vector of cluster labels: if $\gamma_j = 0$, then $\mu^j_{\lambda_t}$, the j-th component of μ_{λ_t}, are the same for all $\lambda_t \in \{1, \ldots, G\}$, i.e., $\mu^j_1 = \cdots = \mu^j_G$; otherwise, they are different. Defining $y^*_t = A(L)y_t$, we have

$$y^*_t \sim N(A(1)\mu_{\lambda_t}, \Sigma_{\lambda_t}),$$

where $A(1) = I - \sum_{i=1}^p A_i$. Let $y^* = (y^*_1, \ldots, y^*_T)$.

The Prior

For the variable selection indicator, we assume that its elements are i.i.d. Bernoulli random variables, i.e., $\gamma_j \sim Ber(\psi)$, with joint density

$$p(\gamma) = \prod_{i=1}^{n} \psi^{\gamma_i}(1 - \psi)^{1-\gamma_i},$$

where ψ can be elicited as the proportion of variables expected a priori to exhibit multiple changes in steady states. In our application we set $\psi = 0.5$. The regime labels $\lambda_t, t = 1, \ldots, T$ are i.i.d. discrete random variables with a hierarchical prior given by $\mathbb{P}(\lambda_t = j) = q_j \geq 0$ with $\sum_{i=1}^G q_i = 1$. In turn, $q = (q_1, \ldots, q_G)$ is a random variable with a symmetric Dirichlet prior, $q \sim Dir(\alpha_0, \ldots, \alpha_0)$, where α_0 is set to be 3 in our application.

The prior for $a = vecA' = vec([A_1, \ldots, A_p]')$ is Normal with mean and covariance matrix obtained as follows: we first estimate the VAR coefficients in the time-varying intercept model

$$y_t = \tilde{\mu}_t + \tilde{A}_1 y_{t-1} + \cdots + \tilde{A}_p y_{t-p} + v_t, \quad v_t \sim \mathsf{N}(0, \Omega),$$

with a random walk state equation

$$\tilde{\mu}_t = \tilde{\mu}_{t-1} + w_t, \quad w_t \sim N(0, Q),$$

where Q is a diagonal matrix. Then the prior mean is set to be $\mathbb{E}(\tilde{A}\,|\,y)$ with prior covariance matrix $0.01^2 \times I$, i.e., $a \sim N(a_0, 0.01^2 \times I)$, where $a_0 = vec(\mathbb{E}(\tilde{A}\,|\,y)')$.

Finally, we assume the following natural conjugate priors for $\mu_1, \ldots, \mu_G$ and $\Sigma_1, \ldots, \Sigma_G$:

$$\begin{aligned}
\mu_{(\gamma=0)}|A(1)_{(\gamma=0)}, \Sigma_{(\gamma=0)} &\sim N(\eta_{0(\gamma=0)}, h_0 A(1)^{-1}_{(\gamma=0)} \Sigma_{(\gamma=0)} A(1)^{-1'}_{(\gamma=0)}), \\
\mu_{j,(\gamma=1)}|A(1)_{(\gamma=1)}, \Sigma_{j,(\gamma=1)} &\sim N(\eta_{0(\gamma=1)}, h_1 A(1)^{-1}_{(\gamma=1)} \Sigma_{j,(\gamma=1)} A(1)^{-1'}_{(\gamma=1)}), \\
\Sigma_{(\gamma=0)} &\sim IW(\delta + n - n_\gamma, Q_{0(\gamma=0)}), \\
\Sigma_{(\gamma=1)} &\sim IW(\delta + n_\gamma, Q_{1(\gamma=1)}).
\end{aligned}$$

where $IW(\cdot,\cdot)$ denotes the inverse-Wishart distribution. The degree of freedom parameter, δ, is set to be n, while $\eta_0 = 0$ and $h_0 = h_1 = 10$ to indicate weak prior information. Finally, we take $Q_0 = 1/\tau_0 I$ and $Q_1 = 1/\tau_1 I$, where τ_0 and τ_1 are chosen as suggested in Tadesse et al (2005).

MCMC Algorithm

Our posterior simulator is based on the collapsed sampler proposed in Tadesse et al (2005). Specifically, we sample the parameters and latent variables marginally of $\mu_1, \ldots, \mu_G$ and $\Sigma_1, \ldots, \Sigma_G$. Note that given γ and a, $p(y^*, \lambda|q, \gamma, a)$, the joint distribution of (y^*, λ) marginal of $\mu_1, \ldots, \mu_G$ and $\Sigma_1, \ldots, \Sigma_G$, is available analytically (see equation 7). Using derivations similar to Tadesse et al (2005), page 605, it can be shown that

$$\begin{aligned}
p(y^*, \lambda|q, \gamma, a) = \pi^{-Tn/2} \prod_{k=1}^{G} \left\{ K_{k(\gamma=1)} |Q_{1(\gamma=1)}|^{(\delta+n_\gamma-1)/2} |Q_{1(\gamma=1)} + S_{k(\gamma=1)}|^{-(T_k+\delta+n_\gamma-1)/2} \right\} \\
\times H_{(\gamma=0)} |Q_{0(\gamma=0)}|^{(\delta+n-n_\gamma-1)/2} |Q_{0(\gamma=0)} + S_{0(\gamma=0)}|^{-(T+\delta+n-n_\gamma-1)/2},
\end{aligned} \tag{7}$$

where

$$K_{k(\gamma=1)} = q_k^{T_k}(1+h_1T_k)^{-n_\gamma/2}\prod_{j=1}^{n_\gamma}\frac{\Gamma((T_k+\delta+n_\gamma-j)/2)}{\Gamma((\delta+n_\gamma-j)/2)},$$

$$H_{(\gamma=0)} = (1+h_0T)^{-(n-n_\gamma)/2}\prod_{j=1}^{n-n_\gamma}\frac{\Gamma((T+\delta+n-n_\gamma-j)/2)}{\Gamma((\delta+n-n_\gamma-j)/2)},$$

$$\begin{aligned}S_{k(\gamma=1)} &= \sum_{t\in C_j}(y^*_{t(\gamma=1)}-\bar{y}^*_{t(\gamma=1)})(y^*_{t(\gamma=1)}-\bar{y}^*_{t(\gamma=1)})' \\ &\quad + \frac{T_k}{h_1T_k+1}(A(1)_{(\gamma=1)}\eta_{0(\gamma=1)}-\bar{y}^*_{t(\gamma=1)})(A(1)_{(\gamma=1)}\eta_{0(\gamma=1)}-\bar{y}^*_{t(\gamma=1)})',\end{aligned}$$

$$\begin{aligned}S_{0(\gamma=0)} &= \sum_{t=1}^{T}(y^*_{t(\gamma=0)}-\bar{y}^*_{t(\gamma=0)})(y^*_{t(\gamma=0)}-\bar{y}^*_{t(\gamma=0)})' \\ &\quad + \frac{T}{h_0T+1}(A(1)_{(\gamma=0)}\eta_{0(\gamma=0)}-\bar{y}^*_{t(\gamma=0)})(A(1)_{(\gamma=0)}\eta_{0(\gamma=0)}-\bar{y}^*_{t(\gamma=0)})',\end{aligned}$$

$n_\gamma = \sum_{i=1}^{n}\gamma_i$, T_k is the number of λ_t that are equal to k, i.e., $T_k = \sum_{t=1}^{T}1(\lambda_t=k)$, $\bar{y}^*_{t(\gamma=1)}$ and $\bar{y}^*_{t(\gamma=0)}$ are the sample means of $y^*_{t(\gamma=1)}$ and $y^*_{t(\gamma=0)}$ respectively.

The posterior simulator now consists of the following four steps:
(1) sample $\gamma|y,\lambda,q,a$;
(2) sample $\lambda|y,\gamma,q,a$;
(3) sample $q|y,\lambda,\gamma,a$;
(4) sample $a|y,\lambda,\gamma,q$;

The details of Steps (1)-(2) are given in Tadesse et al (2005) equations (10)-(12) and the discussion in their subsections 5.1–5.2, with the likelihood given in (7). Step 3 is a simple Gibbs step as (conditional on λ), q has a Dirichlet distribution: $q|y,\lambda,\gamma,a \sim Dir(\alpha_1,\ldots,\alpha_G)$, where $\alpha_k = \alpha_0 + T_k$ and T_k is the number of λ_t that are equal to k. To avoid the label-switching problem, we impose the restriction that $q_1 \leq \cdots \leq q_G$. Such a draw can be obtained, for example, by rejection sampling. To implement Step 4, recall that the prior for a is $a \sim N(a_0, 0.01^2 \times I)$. Since the prior is tight, we can use $p(a)$ as the proposal density in an independence-chain sampler. Specifically, given the current draw a, a candidate draw a^c is generated from $N(a_0, 0.01^2 \times I)$. If the characteristic roots of the companion matrix associated with a^c are

all within the unit circle, we accept a^c with probability $\min\{\alpha_{\text{MH}}, 1\}$, where

$$\alpha_{\text{MH}} = \frac{p(y^{*c}, \lambda | q, \gamma, a^c)}{p(y^*, \lambda | q, \gamma, a)},$$

$y_t^{*c} = A^c(L) y_t$ and $A^c(L)$ is the lag polynomial associated with a^c; otherwise, retain the draw a. Since a is a high-dimensional vector, in our implementation we divide a into several blocks. In particular, each row of A constitutes a block, and each block is sampled sequentially.

The BIC used in the paper is defined as $\text{BIC} = -2 \ln \widehat{L} + m \ln T$, where $\widehat{L}$ is the maximized likelihood value and m is the number of parameters in the model. For $G = 1$, $\widehat{L}$ is obtained as follows: first the likelihood function is evaluated (in this model we don't have any latent variables so the likelihood value can be easily obtained) at each posterior draw, and $\widehat{L}$ is set to be the maximum value. For $G > 1$ cases, $\widehat{L}$ is approximated by the average of the complete data likelihood values evaluated at each posterior draw.

Posterior results in the empirical application are based on 50000 posterior draws, following a burn-in period of 1000. Results pass standard checks of MCMC convergence.