Computer Tutorial 2: Hierarchical Priors for Bayesian VARs

In this exercise sheet, I provide a series of questions and answers relating to hierarchical priors in VARs. Exercise 1 is taken from the new edition of Bayesian Econometric Methods and includes both theoretical derivations and an empirical application. As with Computer Tutorial 1, I am not expecting you to go through the theoretical derivations and proofs in detail. Rather, the computer session will focus on the computational part of the exercise (part b). Code which does these exercises in provided. In the computer lab, I suggest you experiment with these codes (e.g. try different priors, lag lengths, etc.) to familiarize yourself with Bayesian programming in VARs.

Exercise 1 (Stochastic Search Variable Selection in VARs)

SSVS methods are a potentially useful way of surmounting over-parameterization and over-fitting concerns which arise when the number of potential explanatory variables is large relative to the number of observations. VARs, where each of n equations contains p lags of each of n dependent variables, also have many potential explanatory variables. And the error covariance matrix, Σ, contains $\frac{n(n-1)}{2}$ free elements. Thus, with large VARs, involving dozens or more dependent variables, there are a huge number of parameters to estimate. Concerns about over-parameterization and over-fitting are likely to be especially severe with VARs. In light of these concerns, you are asked to:

(a) Derive a Gibbs sampling algorithm which allows for Bayesian estimation of the VAR using a SSVS prior (VAR-SSVS)

(b) Use the data set from Computer Tutorial 1 estimate the VAR-SSVS. Examine how sensitive your findings are to prior hyperparameter choices.

Solution to Exercise 1

(a) This solution sets up the prior and derives the posterior conditionals in the same manner as George, Sun and Ni (2008).

We use notation where β is the $K \times 1$ vector of VAR coefficients and Σ is the $n \times n$ error covariance matrix. We work with the Cholesky decomposition:

$$\Sigma^{-1} = \Psi\Psi'$$

where Ψ is upper-triangular with elements ψ_{ij}. The diagonal elements of the latter matrix are denoted by $\psi = (\psi_{11}, .., \psi_{nn})'$ and the off-diagonal upper-triangular elements by $\eta = (\eta_2', .., \eta_n')'$ where $\eta_j = \left(\psi_{1j}, .., \psi_{j-1,j}\right)'$ for $j = 2, .., n$.

The model is parameterized in terms of β, ψ and η. SSVS can be interpreted as defining a particular hierarchical prior for these parameters. In particular, the prior for the VAR coefficients is

$$\beta|\gamma^{\beta} \sim N\left(0, \underline{D}^{\beta}\right) \tag{1}$$

where $\underline{D}^{\beta}$ is a diagonal matrix with diagonal elements $\underline{d}_i^{\beta}$ for $i = 1, .., K$ and $\gamma^{\beta} = \left(\gamma_1^{\beta}, .., \gamma_k^{\beta}\right)'$. The SSVS nature of this prior arises by having $\gamma_i^{\beta} \in \{0, 1\}$ and defining:

$$\underline{d}_i^{\beta} = \begin{cases} \tau_{0i}^2 \text{ if } \gamma_i^{\beta} = 0 \\ \tau_{1i}^2 \text{ if } \gamma_i^{\beta} = 1 \end{cases}$$

where τ_{0i}^2 is small and τ_{1i}^2 is large.[1] Thus, if $\gamma_i^{\beta} = 0$ the corresponding VAR coefficient is shrunk to be close to zero, whereas if $\gamma_i^{\beta} = 1$ it is estimated in a relatively unrestricted fashion. γ^{β} is a vector of unknown parameters and, thus,

[1] George, Sun and Ni (2008) describe a semi-default automatic approach to objectively choosing these small and large prior variances. In the empirical illustration of part b), we subjectively choose them.

requires a prior. We assume the elements of γ^{β} to be, a priori, independent of one another with

$$\Pr\left(\gamma_i^{\beta}=0\right)=\underline{p}_i \tag{2}$$

and, thus, $\Pr\left(\gamma_i^{\beta}=1\right)=1-\underline{p}_i$

The hierarchical prior for η has the same SSVS form allowing for individual error covariances to be shrunk towards zero or estimated in an unrestricted fashion. In particular, we assume, for $j=2,..,n$,

$$\eta_j|\gamma_j^{\eta}\sim N\left(0,\underline{D}_j^{\eta}\right) \tag{3}$$

where $\underline{D}_j^{\eta}=diag\left(\underline{d}_{1j}^{\eta},..,\underline{d}_{j-1,j}^{\eta}\right)$. The vector of SSVS indicator variables, $\gamma_{ij}^{\eta}\in\{0,1\}$, is defined as $\gamma^{\eta}=\left(\gamma_2^{\eta},..,\gamma_n^{\eta}\right)'$ where $\gamma_j^{\eta}=\left(\gamma_{1j}^{\eta},..,\gamma_{j-1,j}^{\eta}\right)'$. Next we assume

$$\underline{d}_{ij}^{\eta}=\begin{cases}\kappa_{0ij}^2 \text{ if } \gamma_{ij}^{\eta}=0\\ \kappa_{1ij}^2 \text{ if } \gamma_{ij}^{\eta}=1\end{cases}$$

where κ_{0ij}^2 is small and κ_{1i}^2 is large. We assume the elements of γ^{η} to be, a priori, independent of one another with

$$\Pr\left(\gamma_{ij}^{\eta}=0\right)=\underline{q}_{ij}. \tag{4}$$

Finally, the error variances are assumed to have Gamma priors (independent of one another) and, thus,

$$\psi_{ii}^2\sim G\left(\underline{a}_i,\underline{b}_i\right) \tag{5}$$

for $i=1,..,n$.

We next derive a Gibbs sampler involving the posterior conditionals for β,ψ, η, γ^{β} and γ^{η}.

We do not provide details of the posterior conditional for β since they are given in the solution to Exercise 1 of Computer Session 1 with prior hyperparameters $\beta_0,V_{\boldsymbol{\beta}}$ of that exercise replaced by $0,\underline{D}^{\beta}$ from (1). That is, conditional on γ^{β}, the SSVS prior for β is a Normal prior of exactly the same form as in yesterday's exercise.

Any Bayesian posterior involves multiplying likelihood times prior. To get a conditional posterior for a specific parameter, one fixes all the other parameters and treats the expression as a p.d.f for the specific parameter. If one follows this strategy for the posterior of γ^{β} conditional on all the other parameters things are greatly simplified by noting that γ^{β} does not enter the likelihood. In fact it only appears in (1) and (2). Multiplying the Normal form for the former by the Bernoulli form for the latter yields a conditional posterior of a simple form. To be precise, the assumption that γ_i^{β} is, a priori, independent of γ_j^{β} for $i\neq j$ means that we can draw the SSVS variable selection indicators independently of one another with probabilities (for $i=1,..,k$)

$$\Pr\left(\gamma_i^{\beta}=0|y,\beta,\eta,\psi\right)=\overline{p}_i$$

where

$$\overline{p}_i = \frac{\tau_{0i}^{-1}\exp\left(-\frac{\beta_i^2}{2\tau_{0i}^2}\right)\underline{p}_i}{\tau_{0i}^{-1}\exp\left(-\frac{\beta_i^2}{2\tau_{0i}^2}\right)\underline{p}_i + \tau_{1i}^{-1}\exp\left(-\frac{\beta_i^2}{2\tau_{1i}^2}\right)\left(1-\underline{p}_i\right)}$$

where the denominator of this expression ensures probabilities sum to one. Furthermore, $\Pr\left(\gamma_i^\beta = 1 | y, \beta, \eta, \psi\right) = 1 - \overline{p}_i$.

To obtain the conditional posterior for η, we use notation from the lecture slides and write the VAR in matrix form and let

$$S = (Y - XA)(Y - XA)'$$

be the $n \times n$ sum of squares matrix of the VAR errors with upper-left $j \times j$ submatrix being denoted by S_j and individual elements of S being s_{ij} and $s_j = (s_{1j}, .., s_{j-1,j})'$. Using the properties of the matric-variate Normal distribution, the likelihood function is proportional to

$$|\Psi|^T \exp\left\{tr\left(-\frac{1}{2}\Psi' S \Psi\right)\right\}.$$

Since Ψ is upper-triangular we have $|\Psi|^T = \prod_{i=1}^{n} \psi_{ii}^T$. And, multiplying out the terms inside the trace operator we have

$$\sum_{i=1}^{n} \psi_{ii}^2 v_i + \sum_{j=2}^{n} \left(\eta_j + \psi_{jj} S_{j-1}^{-1} s_j\right)' S_{j-1} \left(\eta_j + \psi_{jj} S_{j-1}^{-1} s_j\right).$$

Putting these two pieces together we obtain a likelihood proportional to

$$\prod_{i=1}^{n} \psi_{ii}^T \exp\left\{-\frac{1}{2}\left[\sum_{i=1}^{n} \psi_{ii}^2 v_i + \sum_{j=2}^{n} \left(\eta_j + \psi_{jj} S_{j-1}^{-1} s_j\right)' S_{j-1} \left(\eta_j + \psi_{jj} S_{j-1}^{-1} s_j\right)\right]\right\}. \tag{6}$$

Note that the likelihood breaks into independent components for each η_j for $j = 2, .., p$ and each has a Normal form. The priors given by (3) are also Normal and independent of one another. It is straightforward to multiply each Normal prior by the Normal likelihood component for η_j to obtain for $j = 2, .., p$

$$\eta_j | y, \beta, \psi \sim N\left(\overline{\mu}_j, \overline{D}_j^\eta\right)$$

where

$$\overline{D}_j^\eta = \left[S_{j-1} + \left(\underline{D}_j^\eta\right)^{-1}\right]^{-1}$$

$$\overline{\mu}_j = -\psi_{jj} \overline{D}_j^\eta s_j.$$

To obtain the posterior conditionals for γ^η we follow the same strategy (and almost identical derivations) as for γ^β and find (for $j = 2, .., n$ and $i = 1.., j-1$)

$$\Pr\left(\gamma_{ij}^\eta = 0 | y, \beta, \eta, \psi\right) = \overline{q}_{ij}$$

where

$$\overline{q}_{ij} = \frac{\kappa_{0ij}^{-1} \exp\left(-\frac{\psi_{ij}^2}{2\kappa_{0ij}^2}\right) \underline{q}_{ij}}{\kappa_{0ij}^{-1} \exp\left(-\frac{\psi_{ij}^2}{2\kappa_{0ij}^2}\right) \underline{q}_{i} + \kappa_{1ij}^{-1} \exp\left(-\frac{\psi_{ij}^2}{2\kappa_{1ij}^2}\right) \left(1 - \underline{q}_{ij}\right)}$$

where the denominator of this expression ensures probabilities sum to one. Furthermore, $\Pr\left(\gamma_{ij}^{\eta} = 1|y, \beta, \eta, \psi\right) = 1 - \overline{q}_{ij}$.

Finally, to obtain draws of ψ, this can be done by using the independent Gamma conditional posteriors for ψ_{jj}^2 for $j = 1, .., n$ which are:

$$\psi_{ii}^2 | y, \beta, \gamma^{\beta} \sim G\left(\overline{a}_i, \overline{b}_i\right)$$

where

$$\begin{aligned}
\overline{a}_i &= \underline{a}_i + \frac{1}{2} \\
\overline{b}_1 &= \underline{b}_1 + \frac{s_{11}}{2} \\
\overline{b}_j &= \underline{b}_j + \frac{1}{2}\left\{ s_{jj} + s_j' \left[S_{j-1} + \left(\underline{D}_j^{\eta}\right)^{-1} \right]^{-1} s_j \right\} \text{ for } j = 2, .., n.
\end{aligned}$$

The proof is standard since it involves multiplying the Gamma prior (5) by the Normal likelihood in (6) and collecting terms involving ψ_{ii}^2.

(b) Matlab code (named VAR_SSVS.m) which answers this question is available on this book's website. The results below set $p = 1$ and include an intercept. Prior hyperparameters are set as $\underline{q}_{ij} = \underline{p}_i = \frac{1}{2}$, $\tau_{0i} = \kappa_{0ij} = 0.01$, $\tau_{1i} = \kappa_{1ij} = 100$ $\underline{a}_i = \underline{b}_i = 0.01$ for all i and j (although you may wish to experiment with other prior hyperparameter values).

The following table, based on 22,000 draws of which the first 2,000 were discarded as the burn-in, presents posterior means, standard deviations and the posterior inclusion probability (i.e. $\Pr\left(\gamma_i^{\beta} = 1|y\right)$ for each $i = 1, .., K$). It can be seen that posterior inclusion probabilities are very high for the first own lags in each of the three equations and their coefficients are similar to the comparable Minnesota prior results in Table 1. However, relative to Table 1, the remainder of the coefficients have been shrunk towards zero.

Table 1: Posterior Results for VAR Coefficients using VAR-SSVS			
	Mean	St. Deviation	Inclusion Probability
	Inflation Equation		
Lag inflation	0.793	0.049	1.000
Lag unemployment	0.010	0.007	3×10^{-4}
Lag interest rate	0.020	0.009	0.015
	Unemployment Equation		
Lag inflation	0.093	0.074	0.595
Lag unemployment	0.969	0.007	1.000
Lag interest rate	0.014	0.010	0.003
	Interest Rate Equation		
Lag inflation	0.139	0.207	0.329
Lag unemployment	0.003	0.009	1×10^{-4}
Lag interest rate	0.965	0.035	1.000

Table 2 is of the same format as Table 1, except it contains posterior results for Σ. The SSVS algorithm does not apply to the error variances, only the covariances which is why there are no inclusion probabilities to be reported for the former. It can be seen that there is one covariance, Σ_{21}, that is being shrunk to zero. The VAR-SSVS is providing strong evidence that Σ_{31} and Σ_{32} are non-zero indicating that the error in the interest rate equation is correlated with the errors of the other two equations.

Table 2: Posterior Results for Σ using VAR-SSVS			
	Mean	St. Deviation	Inclusion Probability
Σ_{11}	0.177	0.018	n.a.
Σ_{21}	-0.001	0.005	0.053
Σ_{31}	0.112	0.028	1.000
Σ_{22}	0.089	0.009	n.a.
Σ_{32}	-0.142	0.023	1.000
Σ_{33}	0.910	0.090	n.a.

Exercise 2 (Dirichlet Laplace Hierarchical Priors in VARs)

The Dirichlet Laplace hierarchical prior was covered in the lecture and I will not provide more details of the statistical theory underlying it. If you want to learn more about the theory, the following paper is an excellent source:

Bhattacharya, A., Pati, D., Pillai, N. S., & Dunson, D. B. (2015). Dirichlet–Laplace priors for optimal shrinkage. Journal of the American Statistical Association, 110(512), 1479-1490.

If you want to see a good empirical application which uses this approach in VARs, then the following paper is a good one:

Kastner, G. and Huber, F. (2017). Sparse Bayesian vector autoregressions in huge dimensions available at https://arxiv.org/abs/1704.03239

I provide code (BVAR_DL.m) which estimates this model on a small data set for you to experiment with.