Computer Tutorial 1: Bayesian VARs

In this exercise sheet, I provide a series of questions and answers relating to two Bayesian VARs. Given that I am providing the answers, what do I expect you to do in the computer sessions? First I am not expecting you to go through the theoretical derivations and proofs (parts a and b of Exercise 1 and part a) of Exercise 2). However, since I do not have time to do proofs in the lectures, I felt I should make them available to you in case you want to see them or have them as a resource for your future Bayesian research. So I suggest you do only a quick reading of this exercise sheet (without worrying about the details of the proofs) before the computer tutorial to get an overview of the model each exercise relates to. Instead focus on the computational part of the each exercise (parts c) and b), respectively, of the two exercises). Code which does these exercises is 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.

Some of the notation I use in this exercise sheet is different than that used in the lecture.

Exercise 1 (The Vector Autoregressive Model with Independent Normal-Wishart Prior)

Suppose $y_t = (y_{1t}, \ldots, y_{nt})'$ is a vector of n dependent variables at time t for $t = 1, \ldots, T$. Then the VAR(p) is defined as:

$$y_t = b + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \epsilon_t, \tag{1}$$

where b is an $n \times 1$ vector of intercepts, $A_1, \ldots, A_p$ are $n \times n$ coefficient matrices and $\epsilon_t \sim N(0, \Sigma)$. In other words, the VAR(p) is simply a multiple-equation regression where the regressors are the lagged dependent variables.[1]

The VAR(p) can be written as:

$$y_t = X_t \beta + \epsilon_t,$$

where $X_t = I_n \otimes [1, y'_{t-1}, \ldots, y'_{t-p}]$ and $\beta = \text{vec}([b,\ A_1,\ \cdots\ A_p]')$. Alternatively, the observations can be stacked over $t = 1, \ldots, T$ so that the VAR(p) can be written as

$$y = X\beta + \epsilon$$

where $\epsilon \sim N(0, I_T \otimes \Sigma)$.

(a) Derive the likelihood function for the VAR.

(b) Assume that the priors for β and Σ are independent of one another and take the form:

$$\beta \sim N(\beta_0, V_{\boldsymbol{\beta}}), \quad \Sigma \sim IW(\nu_0, S_0) \tag{2}$$

where $IW(.,.)$ is the inverse-Wishart distribution. This is often referred to as the independent Normal-Wishart prior. Derive a Gibbs sampler which allows for posterior inference in the VAR.

(c) Use the data set US_macrodata.csv provided which contains US quarterly data on the CPI inflation rate, the unemployment rate and Fed Funds rate from 1959Q2 to 2007Q4—the sample ends at 2007Q4 to avoid the periods when interest rate hits the zero lower bound. Estimate a VAR using this data set and the Gibbs sampler derived in part b). Calculate impulse responses of all variables to a monetary policy shock.

Solution to Exercise 1

(a) Since the VAR(p) is a multivariate regression model, standard Bayesian derivations used with multivariate Normal analysis apply (see Exercise 2.14). In particular, since

$$(y \,|\, \beta, \Sigma) \sim N(X\beta, I_T \otimes \Sigma),$$

the likelihood function is given by:

$$\begin{aligned} p(y \,|\, \beta, \Sigma) &= |2\pi(I_T \otimes \Sigma)|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}(y - X\beta)'(I_T \otimes \Sigma)^{-1}(y - X\beta)} \\ &= (2\pi)^{-\frac{Tn}{2}} |\Sigma|^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2}(y - X\beta)'(I_T \otimes \Sigma^{-1})(y - X\beta)}, \end{aligned} \tag{3}$$

[1] Note that the VAR depends on p initial conditions $y_{-p+1}, \ldots, y_0$. In principle these initial conditions can be modeled explicitly. However, it is common to treat them as fixed (i.e. the first p observations used as initial conditions and the estimation sample begins with the $(p+1)^{th}$ observation). We adopt this strategy in these exercises.

where the second equality holds because $|I_T \otimes \Sigma| = |\Sigma|^T$ and $(I_T \otimes \Sigma)^{-1} = I_T \otimes \Sigma^{-1}$. Note that the likelihood can also be written as

$$p(y \,|\, \beta, \Sigma) = (2\pi)^{-\frac{Tn}{2}} |\Sigma|^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{t=1}^{T}(y_t - X_t\beta)'\Sigma^{-1}(y_t - X_t\beta)}. \tag{4}$$

(b) We derive a Gibbs sampler for the VAR(p) with likelihood given in (3) and priors given in (2). Specifically, we derive the two conditional densities $p(\beta \,|\, y, \Sigma)$ and $p(\Sigma \,|\, y, \beta)$.

The first step is straightforward, as standard linear regression results apply (see Chapter 10 of Bayesian Econometric Methods). In fact, we have

$$(\beta \,|\, y, \Sigma) \sim N(\hat{\beta}, K_\beta^{-1}),$$

where

$$K_\beta = V_\beta^{-1} + X'(I_T \otimes \Sigma^{-1})X, \quad \hat{\beta} = K_\beta^{-1}\left(V_\beta^{-1}\beta_0 + X'(I_T \otimes \Sigma^{-1})y\right),$$

and we have used the result $(I_T \otimes \Sigma)^{-1} = I_T \otimes \Sigma^{-1}$.

Next, we derive the conditional density $p(\Sigma \,|\, y, \beta)$. We will use the fact that, for conformable matrices A, B, C, we have

$$\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB).$$

Combining the likelihood in (3) with the inverse-Wishart prior for Σ, we obtain

$$\begin{aligned}
p(\Sigma \,|\, y, \beta) &\propto p(y \,|\, \beta, \Sigma)p(\Sigma) \\
&\propto |\Sigma|^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{t=1}^{T}(y_t - X_t\beta)'\Sigma^{-1}(y_t - X_t\beta)} \times |\Sigma|^{-\frac{\nu_0+n+1}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}(S_0\Sigma^{-1})} \\
&\propto |\Sigma|^{-\frac{\nu_0+n+T+1}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}(S_0\Sigma^{-1})} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left[\sum_{t=1}^{T}(y_t - X_t\beta)(y_t - X_t\beta)'\Sigma^{-1}\right]} \\
&\propto |\Sigma|^{-\frac{\nu_0+n+T+1}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left[\left(S_0 + \sum_{t=1}^{T}(y_t - X_t\beta)(y_t - X_t\beta)'\right)\boldsymbol{\Sigma}^{-1}\right]},
\end{aligned}$$

which is the kernel of an inverse-Wishart density (type "inverse-Wishart" into Wikipedia to see this). Thus, we have

$$(\Sigma \,|\, y, \beta) \sim IW\left(\nu_0 + T, S_0 + \sum_{t=1}^{T}(y_t - X_t\beta)(y_t - X_t\beta)'\right).$$

We summarize the Gibbs sampler as follows

Algorithm 1 ***(Gibbs Sampler for the VAR(p)).***

Pick some initial values $\beta^{(0)} = \mathrm{c}_0$ *and* $\Sigma^{(0)} = \mathrm{C}_0 > 0$. *Then, repeat the following steps from* $\mathrm{r} = 1$ *to* R*:*

1. *Draw* $\beta^{(r)} \sim p(\beta \,|\, y, \Sigma^{(r-1)})$ *(multivariate Normal).*

2. *Draw* $\Sigma^{(\mathrm{r})} \sim p(\Sigma \,|\, y, \beta^{(r)})$ *(inverse-Wishart).*

(c) Matlab code which answers this question and produces the figures below is provided. Empirical results below use a VAR(2) and sets $\beta_0 = 0, V_{\boldsymbol{\beta}} = I$ (except for the diagonal elements of $V_{\boldsymbol{\beta}}$ which correspond to the intercepts in each equation which are set to 10), $\nu_0 = 6, S_0 = I$. In order to identify the monetary policy shocks we order the interest rate last and treat it as the monetary policy instrument.

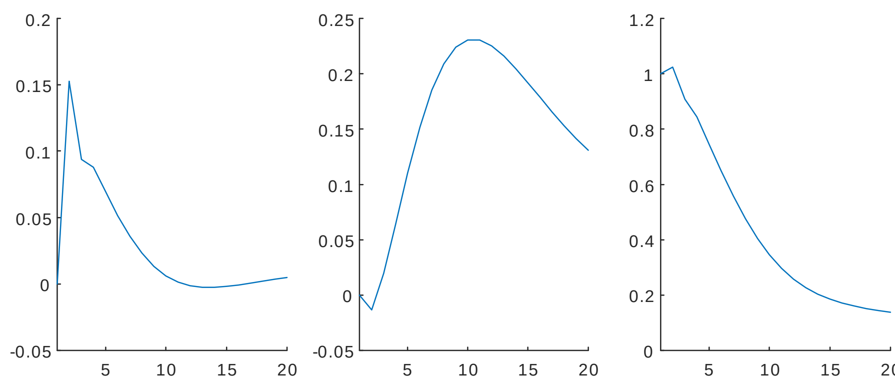
We first implement the Gibbs sampler described in part b). Then, given the posterior draws of β and Σ, we compute the impulse-response functions of the three variables to a 100-basis-point monetary policy shock. Thus, the MATLAB code contains several scripts. The main script is VAR.m. It loads the dataset, constructs the regression matrix X using the script SURform2.m.

For each of the posterior draws of β and Σ, we use the script construct_IR.m to compute the impulse-response functions of the three variables to a 100-basis-point monetary policy shock. More specifically, we consider two alternative paths: in one a 100-basis-point monetary policy shock hits the system, and in the other this shock is absent. We then let the two systems evolve according to the VAR(p) written as the regression

$$y_t = X_t \beta_t + C\tilde{\epsilon}_t, \quad \tilde{\epsilon}_t \sim N(0, I_3),$$

for $t = 1, \ldots, n_{\mathrm{hz}}$, where n_{hz} denotes the number of horizons and C is the Cholesky factor of Σ. Each impulse-response function is then the difference between these two paths.

The impulse-response functions of inflation, unemployment and interest rate to a 100-basis-point monetary policy shock are given in the figure below.



Impulse-response functions of inflation (left panel), unemployment (middle panel) and interest rate (right panel) to a 100-basis-point monetary policy shock.

By construction, the interest rate increases by 1% on impact, and it gradually goes back to zero over the next 20 quarters. The impact on unemployment is delayed and positive—the unemployment rate rises slowly and reaches a peak of about 0.25% after 3 years on impact.

In contrast, the inflation rate rises shortly after impact, and the effect is positive. The literature often refers to this empirical finding as the price puzzle—a contractionary monetary policy shock should dampen rather than increase inflation. Some argue that this reflects misspecification—some crucial variables are omitted, and both interest rate and inflation are responding to these omitted variables. Alternatively, Primiceri (2005) argues that the monetary policy shocks identified in this context should be interpreted as non-systematic policy actions that capture both policy mistakes and interest rate movements that are responses to variables other than inflation and unemployment.

Exercise 2 (The VAR with Minnesota Prior)

For this exercise, we use the notation and data set of Exercise 1. The Minnesota prior has traditionally been the most popular prior for Bayesian VARs. It involves a simplification in that Σ is assumed to be a diagonal matrix and replaced with an estimate, $\widehat{\Sigma}$, where $\widehat{\sigma}_{ii} = s_i^2$ (where s_i^2 is the standard OLS estimate of the error variance using an AR model for the i^{th} variable and $\widehat{\sigma}_{ii}$ is the ii^{th} element of $\widehat{\Sigma}$). To explain the prior for the vector of VAR(p) coefficients, β, note that some elements correspond to own lags (e.g. in the equation for the first variable in the VAR there are lags of the first variable), some elements correspond to other lags (e.g. in the equation for the first variable in the VAR there are lags of the second and third variables) and some to exogenous or deterministic variables (e.g. the intercept). The Minnesota prior reflects a belief that own lags are more likely to be important explanatory variables than other lags and more recent lags are likely to be more important than those in the more distant past.

The Minnesota prior assumes $\beta \sim N(\beta_0, V_{\boldsymbol{\beta}})$. Consider first, the prior mean, β_0. When using stationary variables which exhibit little persistence (e.g. the growth rates of macroeconomic variables) it is common to simply set $\beta_0 = 0$. When working with nonstationary data (e.g. the logs of macroeconomic variables) it is common to set $\beta_0 = 0$ except for the elements corresponding to the first own lag of the dependent variable in each equation. These elements are set to one. Next consider the prior covariance matrix, $V_{\boldsymbol{\beta}}$, which is assumed to be a diagonal matrix. The Minnesota prior sets the diagonal elements of $V_{\boldsymbol{\beta}}$ as follows:

- $\frac{\underline{a}_1}{r^2}$ for coefficients on own lag r for $r = 1, .., p$
- $\frac{\underline{a}_2 \widehat{\sigma}_{ii}}{r^2 \widehat{\sigma}_{jj}}$ for coefficients on lag r of variable $j \neq i$ for $r = 1, .., p$ in equation i
- $\underline{a}_3 \widehat{\sigma}_{ii}$ for coefficients on deterministic or exogenous variables in equation i.

The Minnesota prior thus turns a complicated prior elicitation procedure into a much simpler one which requires the research to select three scalar prior hyperparameters, $\underline{a}_1, \underline{a}_2$ and $\underline{a}_3$. Typically the researcher sets $\underline{a}_1 > \underline{a}_2$. In cases where the number of deterministic terms is small (e.g. if the only such terms are intercepts) then $\underline{a}_3$ is set to a large number, reflecting relatively non-informative prior beliefs.

(a) Derive the posterior for the VAR parameters using the Minnesota prior. Discuss what computational methods can be used to carry out Bayesian inference with this posterior.

(b) Use the data set described in Exercise 1. Estimate a VAR with the Minnesota prior using this data set.

Solution to Exercise 2

(a) This derivation involves working out the posterior for β conditional on $\Sigma = \widehat{\Sigma}$. But an identical derivation was done in Exercise 1 as part of the derivation of the Gibbs sampler in that exercise. Thus, we can simply use our previous derivation to say:

$$(\beta \mid y) \sim N(\hat{\beta}, K_\beta^{-1}),$$

where

$$K_\beta = V_\beta^{-1} + X'(I_T \otimes \widehat{\Sigma}^{-1})X, \quad \hat{\beta} = K_\beta^{-1}\left(V_\beta^{-1}\beta_0 + X'(I_T \otimes \widehat{\Sigma}^{-1})y\right),$$

where the Minnesota prior forms for β_0 and V_β are used. This implies that posterior inference can be done analytically. That is, the posterior mean of the VAR coefficients is $\hat{\beta}$ and the posterior covariance matrix is K_β^{-1}. The properties of the Normal distribution can be used to calculate other features such as credible intervals.

Thus, there is no need to use any posterior simulation method with the Minnesota prior unless interest centers on nonlinear functions of β such as impulse response functions. Monte Carlo integration can be used to carry out posterior inference on features such as impulse response functions. That is, draws can be taken from the Normal posterior for β and each transformed to produce an impulse response function. Averages of these draws will converge to the posterior mean of the impulse response function.

(b) Matlab code (named bvar_minnesota.m) which answers this question is provided. The results below set $p = 1$ and $\underline{a}_1 = 0.5, \underline{a}_2 = 0.25$ and $\underline{a}_3 = 100$ and include an intercept. We have also set $\beta_0 = 0$ (although the user may wish to experiment with other values as the variables do display a fairly high degree of persistence). The following table presents estimates of the VAR coefficients along with their posterior standard deviations. Each column of the table presents results for one of the three equations in the VAR. If we use the rule of thumb which says a variable has important explanatory power if its posterior mean is more than two posterior standard deviations from zero, then the first own lag is important in every equation. Beyond this, it is only the unemployment equation which has any important explanatory variables.

Table 21.1: Posterior Results using Minnesota Prior for the VAR
Posterior mean (stand. dev. in parentheses)

	Inflation	Unemployment	Interest Rate
Lag inflation	0.701 (0.057)	0.088 (0.040)	0.212 (0.128)
Lag unemployment	-0.028 (0.022)	0.953 (0.016)	-0.050 (0.050)
Lag interest rate	0.038 (0.013)	0.023 (0.009)	0.927 (0.303)