A general model. The Multivariate Normal linear model is

$$\begin{aligned} Y_t' &= X_t'B + u_t' \\ u_t' &\sim N(\mathbf{0}_{1\times n}, \Sigma_u), \end{aligned} \tag{1}$$

where Y_t is a $1 \times n$ row vector. This model nests within it many other models, including both VARs and VECMs, as well as seemingly-unrelated regressions (SURs). For a VAR of order p,

$$X_t' = \left[1, \{Y_{t-j}'\}_{j=1}^{p}\right], \tag{2}$$

while, for a VECM,

$$X_t' = \left[1, Y_{t-1}, \{\Delta Y_{t-j}'\}_{j=1}^{p-1}\right], \tag{3}$$

where $\Delta Y_t = Y_t - Y_{t-1}$. The particular variables in X_t aren't particularly important for computing the forms of the posteriors, except for specifying an appropriate prior, so we'll be ignoring that until we look at specific priors. We'll just say that X_t' is an $1 \times m$ row vector instead.

Priors and likelihoods for the general model. The conjugate priors we'll use for this model are Matrix Normal and Inverse Wishart:

$$\begin{aligned} B \mid \Sigma_u &\sim MN(\overline{B}, \Omega, \Sigma_u) \\ \Sigma_u &\sim IW(\Psi, d), \end{aligned} \tag{4}$$

where Ψ is an $n \times n$ matrix and Ω is an $m \times m$ matrix.[1] Both have very standard forms. The Matrix Normal may be slightly less familar, but it's closely related to the Multivariate Normal distribution, as, we'll see later.

The PDF of a $B \mid \Sigma_u$, given that it is distributed $MN(\overline{B}, \Omega, \Sigma_u)$, is

$$\pi(B \mid \Sigma_u) = (2\pi)^{-\frac{nm}{2}} |\Sigma_u|^{-\frac{m}{2}} |\Omega|^{-\frac{n}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}(B-\overline{B})'\Omega^{-1}(B-\overline{B})\right)\right), \tag{5}$$

[1] For VARs and VECMs of order p with n variables, $m = np + 1$.

while the PDF of Σ_u, given that it is distributed $\mathrm{IW}(\Psi, d)$, is

$$\pi(\Sigma_u) = 2^{-\frac{dn}{2}} \frac{|\Psi|^{\frac{d}{2}}}{\Gamma_n\left(\frac{d}{2}\right)} |\Sigma_u|^{-\frac{d+n+1}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}\Psi\right)\right). \tag{6}$$

Let $\mathbf{X}$ and $\mathbf{Y}$ be the stacked values of X_t' and Y_t' for $t = 0$ to $T - 1$, respectively. Then the distribution of $\mathbf{Y}$ is

$$\mathbf{Y} \mid \mathbf{X}, B, \Sigma_u \sim \mathrm{MN}(\mathbf{X}B, \mathbf{I}_T, \Sigma_u), \tag{7}$$

which means it has the likelihood

$$P(\mathbf{Y} \mid \mathbf{X}, B, \Sigma_u) = (2\pi)^{-\frac{nT}{2}} |\Sigma_u|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}(\mathbf{Y} - \mathbf{X}B)'(\mathbf{Y} - \mathbf{X}B)\right)\right). \tag{8}$$

Distribution of $B \mid \Sigma_u, \mathbf{X}, \mathbf{Y}$. By Bayes rule, we know

$$P(B \mid \Sigma_u, \mathbf{X}, \mathbf{Y}) \propto P(\mathbf{Y} \mid \mathbf{X}, B, \Sigma_u)\,\pi(B \mid \Sigma_u) \tag{9}$$

To find the posterior, we first compute the product

$$\begin{aligned} P(\mathbf{Y} \mid \mathbf{X}, B, \Sigma_u)\,\pi(B \mid \Sigma_u) = {} & (2\pi)^{-\frac{n(m+T)}{2}} |\Sigma_u|^{-\frac{m+T}{2}} |\Omega|^{-\frac{n}{2}} \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}(\mathbf{Y}'\mathbf{Y} - B'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}B + B'\mathbf{X}'\mathbf{X}B)\right)\right) \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}(\overline{B}'\Omega^{-1}\overline{B} - B'\Omega^{-1}\overline{B} - \overline{B}'\Omega^{-1}B + B'\Omega^{-1}B)\right)\right) \end{aligned} \tag{10}$$

$$\begin{aligned} = {} & (2\pi)^{-\frac{n(m+T)}{2}} |\Sigma_u|^{-\frac{m+T}{2}} |\Omega|^{-\frac{n}{2}} \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}\left(\mathbf{Y}'\mathbf{Y} + \overline{B}'\Omega^{-1}\overline{B}\right)\right)\right) \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}\left(-\left(\mathbf{Y}'\mathbf{X} + \overline{B}'\Omega^{-1}\right)B - B'\left(\Omega^{-1}\overline{B} + \mathbf{X}'\mathbf{Y}\right)\right)\right)\right). \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}\left(B'\left(\Omega^{-1} + \mathbf{X}'\mathbf{X}\right)B\right)\right)\right) \end{aligned} \tag{11}$$

Now, define

$$\widehat{\mathrm{B}} = \left(\Omega^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1}\left(\Omega^{-1}\overline{\mathrm{B}} + \mathbf{X}'\mathbf{Y}\right) \tag{12}$$

$$\widehat{\mathrm{V}} = \left(\Omega^{-1} + \mathbf{X}'\mathbf{X}\right)^{-1}, \tag{13}$$

which allows us to rewrite (11) as

$$\begin{aligned} \mathrm{P}(\mathbf{Y} \mid \mathbf{X}, \mathrm{B}, \Sigma_u)\,\pi(\mathrm{B} \mid \Sigma_u) = {} & (2\pi)^{-\frac{n(m+T)}{2}}|\Sigma_u|^{-\frac{m+T}{2}}|\Omega|^{-\frac{n}{2}} \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}\left(\mathbf{Y}'\mathbf{Y} + \overline{\mathrm{B}}'\Omega^{-1}\overline{\mathrm{B}}\right)\right)\right) \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}\left(-\widehat{\mathrm{B}}'\widehat{\mathrm{V}}^{-1}\mathrm{B} - \mathrm{B}'\widehat{\mathrm{V}}^{-1}\widehat{\mathrm{B}} + \mathrm{B}'\widehat{\mathrm{V}}^{-1}\mathrm{B}\right)\right)\right) \end{aligned} . \tag{14}$$

Completing the square, we obtain

$$\begin{aligned} \mathrm{P}(\mathbf{Y} \mid \mathbf{X}, \mathrm{B}, \Sigma_u)\,\pi(\mathrm{B} \mid \Sigma_u) = {} & (2\pi)^{-\frac{n(m+T)}{2}}|\Sigma_u|^{-\frac{m+T}{2}}|\Omega|^{-\frac{n}{2}} \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}\left(\mathbf{Y}'\mathbf{Y} + \overline{\mathrm{B}}'\Omega^{-1}\overline{\mathrm{B}} - \widehat{\mathrm{B}}'\widehat{\mathrm{V}}^{-1}\widehat{\mathrm{B}}\right)\right)\right) \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}\left(\widehat{\mathrm{B}}'\widehat{\mathrm{V}}^{-1}\widehat{\mathrm{B}} - \widehat{\mathrm{B}}'\widehat{\mathrm{V}}^{-1}\mathrm{B} - \mathrm{B}'\widehat{\mathrm{V}}^{-1}\widehat{\mathrm{B}} + \mathrm{B}'\widehat{\mathrm{V}}^{-1}\mathrm{B}\right)\right)\right) \end{aligned} \tag{15}$$

$$\begin{aligned} = {} & (2\pi)^{-\frac{n(m+T)}{2}}|\Sigma_u|^{-\frac{m+T}{2}}|\Omega|^{-\frac{n}{2}} \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}\left(\mathbf{Y}'\mathbf{Y} + \overline{\mathrm{B}}'\Omega^{-1}\overline{\mathrm{B}} - \widehat{\mathrm{B}}'\widehat{\mathrm{V}}^{-1}\widehat{\mathrm{B}}\right)\right)\right) \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}\left(\mathrm{B} - \widehat{\mathrm{B}}\right)'\widehat{\mathrm{V}}^{-1}\left(\mathrm{B} - \widehat{\mathrm{B}}\right)\right)\right). \end{aligned} \tag{16}$$

We notice now that the third term is proportional to a Matrix Normal PDF for the distribution $\mathrm{MN}(\widehat{\mathrm{B}}, \widehat{\mathrm{V}}, \Sigma_u)$. We, therefore, conclude that the the posterior for the coefficients is

$$\mathrm{B} \mid \Sigma_u, \mathbf{X}, \mathbf{Y} \sim \mathrm{MN}(\widehat{\mathrm{B}}, \widehat{\mathrm{V}}, \Sigma_u).$$

We also note that

$$\begin{aligned} P(\mathbf{Y} \mid \mathbf{X}, B, \Sigma_u)\pi(B \mid \Sigma_u) = {} & (2\pi)^{-\frac{nT}{2}}|\Sigma_u|^{-\frac{T}{2}}|\widehat{V}|^{\frac{n}{2}}|\Omega|^{-\frac{n}{2}} \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}\left(\mathbf{Y}'\mathbf{Y} + \overline{B}'\Omega^{-1}\overline{B} - \widehat{B}'\widehat{V}^{-1}\widehat{B}\right)\right)\right), \\ & \times P(B \mid \Sigma_u, \mathbf{X}, \mathbf{Y}) \end{aligned} \tag{17}$$

which will be useful in computing the next posterior.

Distribution of $\Sigma_u \mid \mathbf{X}, \mathbf{Y}$. Turn to the posterior for Σ_u. By Bayes rule, again, we know that

$$P(B, \Sigma_u \mid \mathbf{X}, \mathbf{Y}) \propto P(\mathbf{Y} \mid \mathbf{X}, B, \Sigma_u)\pi(B \mid \Sigma_u)\pi(\Sigma_u) \tag{18}$$

and, by the Law of Total Probability,

$$P(\Sigma_u \mid \mathbf{X}, \mathbf{Y}) = \int P(B, \Sigma_u \mid \mathbf{X}, \mathbf{Y})dB \propto \left(\int P(\mathbf{Y} \mid \mathbf{X}, B, \Sigma_u)\pi(B \mid \Sigma_u)dB\right)\pi(\Sigma_u). \tag{19}$$

From equation (17), we know

$$\begin{aligned} \int P(\mathbf{Y} \mid \mathbf{X}, B, \Sigma_u)\pi(B \mid \Sigma_u)dB = {} & (2\pi)^{-\frac{nT}{2}}|\Sigma_u|^{-\frac{T}{2}}|\widehat{V}|^{\frac{n}{2}}|\Omega|^{-\frac{n}{2}} \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}\left(\mathbf{Y}'\mathbf{Y} + \overline{B}'\Omega^{-1}\overline{B} - \widehat{B}'\widehat{V}^{-1}\widehat{B}\right)\right)\right) \\ & \times \underbrace{\int P(B \mid \Sigma_u, \mathbf{X}, \mathbf{Y})dB}_{=1}, \end{aligned} \tag{20}$$

so

$$\begin{aligned} \left(\int P(\mathbf{Y} \mid \mathbf{X}, B, \Sigma_u)\pi(B \mid \Sigma_u)dB\right)\pi(\Sigma_u) = {} & 2^{-\frac{dn}{2}}\frac{|\Psi|^{\frac{d}{2}}}{\Gamma_n\left(\frac{d}{2}\right)}(2\pi)^{-\frac{nT}{2}}|\Omega|^{-\frac{n}{2}}|\widehat{V}|^{\frac{n}{2}}|\Sigma_u|^{-\frac{T+d+n+1}{2}} \\ & \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma_u^{-1}\left(\mathbf{Y}'\mathbf{Y} + \overline{B}'\Omega^{-1}\overline{B} - \widehat{B}'\widehat{V}^{-1}\widehat{B} + \Psi\right)\right)\right). \end{aligned} \tag{21}$$

If we define

$$\widehat{S} = \mathbf{Y}'\mathbf{Y} + \overline{B}'\Omega^{-1}\overline{B} - \widehat{B}'\left(\Omega^{-1} + \mathbf{X}'\mathbf{X}\right)\widehat{B} + \Psi, \tag{22}$$

we notice that this is proportional to the PDF for the distribution $IW(\widehat{S}, T+d)$. We conclude that the posterior for the covariance matrix is

$$\Sigma_u \mid \mathbf{X}, \mathbf{Y} \sim IW(\widehat{S}, T+d). \tag{23}$$

Moreover, we have (after pulling out appropriate constant terms)

$$\left(\int P(\mathbf{Y} \mid \mathbf{X}, B, \Sigma_u)\,\pi(B \mid \Sigma_u)\,dB\right)\pi(\Sigma_u) = \pi^{-\frac{nT}{2}}\frac{\Gamma_n\left(\frac{T+d}{2}\right)}{\Gamma_n\left(\frac{d}{2}\right)}|\Psi|^{\frac{d}{2}}|\Omega|^{-\frac{n}{2}}|\widehat{V}|^{\frac{n}{2}}|\widehat{S}|^{-\frac{T+d}{2}}P(\Sigma_u \mid \mathbf{X}, \mathbf{Y}). \tag{24}$$

While the representation of $\widehat{S}$ we have above is particularly efficient, the following is another form we can use:

$$\widehat{S} = (\mathbf{Y} - \mathbf{X}\widehat{B})'(\mathbf{Y} - \mathbf{X}\widehat{B}) + (\widehat{B} - \overline{B})'\Omega^{-1}(\widehat{B} - \overline{B}) + \Psi. \tag{25}$$

This is a particularly intuitive form, as it allows us to write the posterior mode of Σ_u, which is

$$\begin{aligned}\frac{\widehat{S}}{T+d+n+1} = {} & \frac{T-m}{T+d+n+1}\frac{(\mathbf{Y} - \mathbf{X}\widehat{B})'(\mathbf{Y} - \mathbf{X}\widehat{B})}{T-m} \\ & + \frac{m}{T+d+n+1}\frac{(\widehat{B} - \overline{B})'\Omega^{-1}(\widehat{B} - \overline{B})}{m}, \\ & + \frac{d+n+1}{T+d+n+1}\frac{\Psi}{d+n+1}\end{aligned} \tag{26}$$

as a weighted average of three constituent estimates of the variance: the sample error covariance, the covariance of deviations in coefficients from the prior mean, and the prior mean from the Inverse Wishart prior.

The marginal likelihood. The marginal distribution is equal to

$$\begin{aligned} P(\mathbf{Y} \mid \mathbf{X}) &= \int \left(\int P(\mathbf{Y} \mid \mathbf{X}, B, \Sigma_u)\, \pi(B \mid \Sigma_u)\, dB \right) \pi(\Sigma_u)\, d\Sigma_u & (27) \\ &= \pi^{-\frac{nT}{2}} \frac{\Gamma_n\left(\frac{T+d}{2}\right)}{\Gamma_n\left(\frac{d}{2}\right)} |\Psi|^{\frac{d}{2}} |\Omega|^{-\frac{n}{2}} |\widehat{V}|^{\frac{n}{2}} |\widehat{S}|^{-\frac{T+d}{2}} \underbrace{\int P(\Sigma_u \mid \mathbf{X}, \mathbf{Y}) d\Sigma_u}_{=1} & (28) \\ &= \pi^{-\frac{nT}{2}} \frac{\Gamma_n\left(\frac{T+d}{2}\right)}{\Gamma_n\left(\frac{d}{2}\right)} |\Psi|^{\frac{d}{2}} |\Omega|^{-\frac{n}{2}} |\widehat{V}|^{\frac{n}{2}} |\widehat{S}|^{-\frac{T+d}{2}}. & (29) \end{aligned}$$

Alternatively, plugging back in our expressions for $\widehat{V}$ and $\widehat{S}$, the marginal distribution may also be written as

$$\begin{aligned} P(\mathbf{Y} \mid \mathbf{X}) = \pi^{-\frac{nT}{2}} \frac{\Gamma_n\left(\frac{T+d}{2}\right)}{\Gamma_n\left(\frac{d}{2}\right)} |\Psi|^{\frac{d}{2}} |\Omega|^{-\frac{n}{2}} \left| \Omega^{-1} + \mathbf{X}'\mathbf{X} \right|^{-\frac{n}{2}} \\ \times \left| \mathbf{Y}'\mathbf{Y} + \overline{B}'\Omega^{-1}\overline{B} - \widehat{B}' \left(\Omega^{-1} + \mathbf{X}'\mathbf{X} \right) \widehat{B} + \Psi \right|^{-\frac{T+d}{2}}. \end{aligned} \quad (30)$$

These are generic results that we use to conduct inference in VAR models of all sizes.

Making draws from the posteriors. Posterior inference on the parameters of the linear model involves draws from the joint distribution $B, \Sigma \mid \mathbf{X}, \mathbf{Y}$, which can be achieved by drawing from $\Sigma_u \mid \mathbf{X}, \mathbf{Y}$ and $B \mid \Sigma_u, \mathbf{X}, \mathbf{Y}$ in sequence. Then, to obtain N draws form the posterior, for each j from 1 to N, we cycle through two steps:

1. Draw $\Sigma_j^{(j)}$ from $IW(\widehat{S}, T + d)$, where $\widehat{S}$ is given by (22) and $\widehat{B}$ is given by (12). This can achieved by the following steps:
   (a) Draw an $n \times (T + d)$ matrix Q, where each entry of Q is i.i.d. $N(0, 1)$;
   (b) Let $\mathrm{Chol}(\widehat{S})$ be the lower triangular matrix in the Cholesky decomposition of $\widehat{S}$.[2] Then $\Sigma_u^{(j)} = \mathrm{Chol}(\widehat{S})(Q'Q)^{-1}\mathrm{Chol}(\widehat{S})'$;
2. Draw $B^{(j)}$ from $MN(\widehat{B}, \widehat{V}, \Sigma_u^{(j)})$, where $\widehat{B}$ is given by (12) and $\widehat{V}$ is given by (13). This can be achieved by the following steps:
   (a) Draw an $m \times n$ matrix W, where each entry of W is i.i.d. $N(0, 1)$;

[2] That is the convention we'll keep going forward: that $\mathrm{Chol}(\cdot)$ refers to the lower-triangular matrix in the decomposition.

(b) Compute $B^{(j)} = \widehat{B} + \text{Chol}(\widehat{V}) W \,\text{Chol}(\Sigma_u^{(j)})'$.

Compared to the implementation where we work with the vectorized coefficients, this offers multiple ways to speed up the calculations. For example, only $W \,\text{Chol}\,(\Sigma_u)'$ needs to be recomputed with each draw. That's cheap as Σ_u is never very large.

An alternative form for the posterior $B \mid \Sigma_u, \mathbf{X}, \mathbf{Y}$**.** In the literature, it's more common to specify the priors as Multivariate Normal than Matrix Normal, though this is less efficient from a computational standpoint. In that case, the prior is

$$\text{vec}(B) \mid \Sigma_u \sim N(\text{vec}(\overline{B}), \Sigma_u \otimes \Omega), \tag{31}$$

by standard results on the relationship between the Matrix Normal and Multivariate Normal distributions. By the same results, the posterior for $\text{vec}(B)$ is

$$\text{vec}(B) \mid \Sigma_u, \widehat{X}, \widehat{Y} \sim N(\text{vec}(\widehat{B}), \Sigma_u \otimes \widehat{V}). \tag{32}$$