Framework

Consider the Normal-Inverse-Wishart (NIW) prior for the pair $(\boldsymbol{B}, \boldsymbol{\Sigma})$, where:

$$\boldsymbol{\Sigma} \sim \mathcal{IW}_n(\nu, \boldsymbol{\Phi})$$

and

$$\boldsymbol{B} \mid \boldsymbol{\Sigma} \sim \mathcal{MN}_{n\times m}(\boldsymbol{\Psi}, \boldsymbol{\Omega}, \boldsymbol{\Sigma})$$

where the density of the Inverse Wishart is:

$$\mathcal{IW}_n(\boldsymbol{\Sigma}; \nu, \boldsymbol{\Phi}) \propto |\boldsymbol{\Sigma}|^{-(\nu+n+1)/2} \exp\left(-\frac{1}{2}\operatorname{tr}(\boldsymbol{\Phi}\boldsymbol{\Sigma}^{-1})\right)$$

and the density of $\boldsymbol{B}$ conditional on $\boldsymbol{\Sigma}$ is:

$$\begin{aligned}
\mathcal{MN}_{n\times m}(\boldsymbol{B}; \boldsymbol{\Psi}, \boldsymbol{\Omega}, \boldsymbol{\Sigma}) &\propto |\boldsymbol{\Sigma}|^{-m/2} \exp\left(-\frac{1}{2}\operatorname{vec}(\boldsymbol{B}-\boldsymbol{\Psi})^\top (\boldsymbol{\Sigma}\otimes\boldsymbol{\Omega})^{-1}\operatorname{vec}(\boldsymbol{B}-\boldsymbol{\Psi})\right)\\
&= |\boldsymbol{\Sigma}|^{-m/2} \exp\left(-\frac{1}{2}\operatorname{tr}\left[(\boldsymbol{B}-\boldsymbol{\Psi})^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{B}-\boldsymbol{\Psi})\boldsymbol{\Sigma}^{-1}\right]\right)\\
&= |\boldsymbol{\Sigma}|^{-m/2} \exp\left(-\frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Omega}^{-1}(\boldsymbol{B}-\boldsymbol{\Psi})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{B}-\boldsymbol{\Psi})\right]\right)
\end{aligned}$$

"Weak" Normal-Inverse-Wishart Prior

Also known as Uligh or Jeffrey Prior

We specify the following parameters:

$$\begin{aligned}
\nu &= 0\\
\boldsymbol{\Phi} &= \mathbf{0}_{n\times n}\\
\boldsymbol{\Psi} &= \mathbf{0}_{n\times m}\\
\boldsymbol{\Omega}^{-1} &= \mathbf{0}_{m\times m} \quad \Rightarrow \quad \boldsymbol{\Omega} = \operatorname{diag}(\infty, \dots, \infty)
\end{aligned}$$

The joint prior density reduces to:

$$\pi(\boldsymbol{B}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(\nu+n+1)/2} = |\boldsymbol{\Sigma}|^{-(n+1)/2}$$

Uniform Prior on join Impulse Responses

Given the parameterization:

$$\begin{aligned}
\nu &= -(m-3) - (n+1)\\
\boldsymbol{\Phi} &= \mathbf{0}_{n\times n}\\
\boldsymbol{\Psi} &= \mathbf{0}_{n\times m}\\
\boldsymbol{\Omega}^{-1} &= \mathbf{0}_{m\times m} \quad \Rightarrow \quad \boldsymbol{\Omega} = \operatorname{diag}(\infty, \dots, \infty)
\end{aligned}$$

the implied prior over $(\boldsymbol{B}, \boldsymbol{\Sigma})$ is:

$$\pi(\boldsymbol{B}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-(\nu+n+1)/2} = |\boldsymbol{\Sigma}|^{(m-3)/2}$$

Minnesota Prior

$$
\begin{aligned}
&\nu = n+2 \\
&\boldsymbol{\Phi} = \mathrm{diag}(\phi_1, \dots, \phi_n), \text{ where } \phi_i = \widehat{\sigma}_i^2(\nu - n - 1) \\
&\boldsymbol{\Psi} = (\boldsymbol{\Psi})_{ij} = \begin{cases} 1 & \text{if } i = j \text{ and } i = \{1, \dots, n\} \\ 0 & \text{otherwise} \end{cases} \\
&\boldsymbol{\Omega} = \begin{bmatrix} \widetilde{\boldsymbol{\Omega}} & \mathbf{0} \\ \mathbf{0} & V_c \end{bmatrix}, \quad \widetilde{\boldsymbol{\Omega}} = \boldsymbol{K} \otimes \boldsymbol{\Phi}^{-1}
\end{aligned}
$$

with $\boldsymbol{K} = \mathrm{diag}\left(\frac{(\nu-n-1)\lambda_1^2}{1^{\lambda_3}}, \dots, \frac{(\nu-n-1)\lambda_1^2}{p^{\lambda_3}}\right)$ and where $\widehat{\sigma}_i^2$ is the residual variance from fitting an AR(p) to variable i. This is related to the fact that

$$\mathbb{E}(\boldsymbol{\Sigma}) = \boldsymbol{\Phi}/(\nu - \mathbf{n} - \mathbf{1}) \text{ hence } \mathbb{E}(\boldsymbol{\Sigma}) = \mathrm{diag}(\widehat{\sigma}^{\mathbf{2}}_{\mathbf{1}}, \dots, \widehat{\sigma}^{\mathbf{2}}_{\mathbf{n}}).$$

$$
\begin{aligned}
\mathbf{K} &= \lambda_1^2 \mathrm{diag}\left(1, \frac{1}{2^{\lambda_3}}, \frac{1}{3^{\lambda_3}}, \dots, \frac{1}{p^{\lambda_3}}\right) \\
\widetilde{\boldsymbol{\Omega}} &= \mathbf{K} \otimes \boldsymbol{\Phi}^{-1} \\
&= \lambda_1^2 \mathrm{diag}\left(1, \frac{1}{2^{\lambda_3}}, \dots, \frac{1}{p^{\lambda_3}}\right) \otimes \mathrm{diag}\left(\frac{1}{\phi_1}, \frac{1}{\phi_2}, \dots, \frac{1}{\phi_n}\right)
\end{aligned}
$$

The resulting matrix $\boldsymbol{\Omega}^{\text{tilde}}$ has block diagonal structure where block (ℓ, ℓ) corresponding to lag ℓ is dimension $n \times n$ and contains:

$$\widetilde{\boldsymbol{\Omega}}_{[\ell]} = \frac{\lambda_1^2}{\ell^{\lambda_3}} \mathrm{diag}\left(\frac{1}{\psi_1}, \frac{1}{\psi_2}, \dots, \frac{1}{\psi_n}\right)$$

Minnesota Prior with Dummy Observations

The Minnesota prior for Vector Autoregression (VAR) models is implemented using the dummy observation approach of Doan, Litterman and Sims (1984). This method transforms the prior beliefs into artificial data points that can be combined with the actual observations to form the posterior distribution.

Dummy Observation Construction

The prior is implemented by constructing artificial data matrices $\mathbf{Y}^*$ and $\mathbf{X}^*$ with dimensions $(m+\lambda_3 n) \times n$ and $(m+\lambda_3 n) \times m$ respectively, where $m = np+1$ is the total number of regressors.

First Lag Prior

For the first lag coefficients, the dummy observations encode the belief that each variable follows a random walk:

$$\mathbf{Y}^*[1:n, 1:n] = \frac{1}{\lambda_1}\text{diag}(\sqrt{\phi_1}, \ldots, \sqrt{\phi_n})$$

$$\mathbf{X}^*[1:n, 1:n] = \frac{1}{\lambda_1}\text{diag}(\sqrt{\phi_1} \ldots, \sqrt{\phi_n})$$

where ψ_i represents the prior variance for variable i and it is $\phi_i = \widehat{\sigma}_i^2$ for $i = 1, \ldots, n$.

Higher-Order Lag Priors

For lags $\ell = 2, 3, \ldots, p$, the prior imposes shrinkage with geometric decay:

$$\mathbf{X}^*[(\ell-1)n + i, (\ell-1)n + i] = \frac{1}{\lambda_1}\sqrt{\phi_i}\ell^{\lambda_2}$$

for $i = 1, 2, \ldots, n$, with all off-diagonal elements set to zero.

Constant Term Prior

The prior for the constant term is set as:

$$\mathbf{X}^*[m, m] = \frac{1}{\sqrt{V_c}}$$

where V_c is the prior variance for the constant.

Error Covariance Prior

To ensure a proper posterior for the error covariance matrix $\mathbf{\Sigma}$, λ_3 additional dummy observations are added:

$$\mathbf{Y}^*[m + (i-1)n + j, j] = \sqrt{\phi_j} \quad \text{for } i = 1, \ldots, \lambda_3,\ j = 1, \ldots, n$$

Posterior Computation

The posterior hyperparameters are computed using standard normal-inverse-Wishart conjugacy:

$$\begin{aligned}
\mathbf{\Omega}^{-1} &= (\mathbf{X}^*)'\mathbf{X}^* \\
\mathbf{\Psi}' &= \mathbf{\Omega}(\mathbf{X}^*)'\mathbf{Y}^* \\
\mathbf{\Phi} &= (\mathbf{Y}^*)'\mathbf{Y}^* - \mathbf{\Psi}\mathbf{\Omega}^{-1}\mathbf{\Psi}' \\
\nu &= \text{size}(\mathbf{Y}^*, 1) - m + 2 = n + 2
\end{aligned}$$

The value of ν is equivalent to have a prior with $\nu = -m + 2$ before seeing the artificial data. The symmetry of $\mathbf{\Psi}$ is enforced numerically: $\mathbf{\Psi} = \frac{1}{2}(\mathbf{\Psi} + \mathbf{\Psi}')$.

Analytical expression for $\boldsymbol{\Omega}^{-1}$

We have that:

$$\boldsymbol{\Omega}^{-1} = (\mathbf{X}^*)'\mathbf{X}^* = \begin{bmatrix} \mathbf{X}^*_{[1:n]} \\ \mathbf{X}^*_{[n+1:2n]} \\ \vdots \\ \mathbf{X}^*_{[(p-1)n+1:pn]} \\ \mathbf{X}^*_{[m]} \\ \mathbf{0}_{n\times m} \end{bmatrix}' \begin{bmatrix} \mathbf{X}^*_{[1:n]} \\ \mathbf{X}^*_{[n+1:2n]} \\ \vdots \\ \mathbf{X}^*_{[(p-1)n+1:pn]} \\ \mathbf{X}^*_{[m]} \\ \mathbf{0}_{n\times m} \end{bmatrix}$$

where the individual blocks are:

$$\begin{aligned} \mathbf{X}^*_{[1:n]} &= \frac{1}{\lambda_1}\text{diag}(\sqrt{\phi_1}, \sqrt{\phi_2}, \ldots, \sqrt{\phi_n}) \quad \text{(First lag)} \\ \mathbf{X}^*_{[(\ell-1)n+1:\ell n]} &= \frac{1}{\lambda_1}\ell^{\lambda_2}\text{diag}(\sqrt{\phi_1}, \sqrt{\phi_2}, \ldots, \sqrt{\phi_n}) \quad \text{(Lag } \ell) \\ \mathbf{X}^*_{[m]} &= \frac{1}{\sqrt{V_c}}\mathbf{e}_m \quad \text{(Constant)} \\ &\mathbf{0}_{n\times m} \text{ (Covariance dummies with) } \lambda_3 = 1 \end{aligned}$$

This yields a block diagonal structure for $\boldsymbol{\Omega}^{-1}$ where block (ℓ, ℓ) corresponding to lag ℓ is dimension $n \times n$ and contains:

$$\boldsymbol{\Omega}^{-1}_{[\ell]} = \frac{\ell^{2\lambda_3}}{\lambda_1^2}\text{diag}\left(\phi_1, \ldots, \phi_n\right)$$

and $\boldsymbol{\Omega}^{-1}_{m,m} = \frac{1}{V_c}$

Analytical Derivation of $\boldsymbol{\Psi}$

We have that:

$$\boldsymbol{\Psi}' = \boldsymbol{\Omega}(\mathbf{X}^*)'\mathbf{Y}^*.$$

but:

$$(\mathbf{X}^*)'\mathbf{Y}^* = \begin{bmatrix} \frac{1}{\lambda_1^2}\text{diag}(\phi_1, \ldots, \phi_n) \\ \mathbf{0}_{(p-1)n\times n} \\ \mathbf{0}_{1\times n} \end{bmatrix}$$

Therefore:

$$\boldsymbol{\Psi}' = \boldsymbol{\Omega} \begin{bmatrix} \frac{1}{\lambda_1^2}\text{diag}(\phi_1, \ldots, \phi_n) \\ \mathbf{0}_{(p-1)n\times n} \\ \mathbf{0}_{1\times n} \end{bmatrix} = \begin{bmatrix} \lambda_1^2\text{diag}(\frac{1}{\phi_1}, \ldots, \frac{1}{\phi_n})\frac{1}{\lambda_1^2}\text{diag}(\phi_1, \ldots, \phi_n) \\ \mathbf{0}_{(p-1)n\times n} \\ \mathbf{0}_{1\times n} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{n\times n} \\ \mathbf{0}_{(p-1)n\times n} \\ \mathbf{0}_{1\times n} \end{bmatrix}$$

Analytical Derivation of $\mathbf{\Phi}$

We need to compute:

$$\mathbf{\Phi} = (\mathbf{Y}^*)'\mathbf{Y}^* - \mathbf{\Psi}\mathbf{\Omega}^{-1}\mathbf{\Psi}'$$

The dummy observation matrix $\mathbf{Y}^*$ has the following block structure:

$$\mathbf{Y}^* = \begin{bmatrix} \mathbf{Y}^*_{[1:n]} \\ \mathbf{Y}^*_{[n+1:2n]} \\ \vdots \\ \mathbf{Y}^*_{[(p-1)n+1:pn]} \\ \mathbf{Y}^*_{[m]} \\ \mathbf{Y}^*_{[m+1:m+n]} \end{bmatrix} = \begin{bmatrix} \frac{1}{\lambda_1}\text{diag}(\sqrt{\phi_1}, \ldots, \sqrt{\phi_n}) \\ \mathbf{0}_{n\times n} \\ \vdots \\ \mathbf{0}_{n\times n} \\ \mathbf{0}_{1\times n} \\ \text{diag}(\sqrt{\phi_1}, \ldots, \sqrt{\phi_n}) \end{bmatrix}$$

Computation of $(\mathbf{Y}^*)'\mathbf{Y}^*$

$$\begin{aligned}(\mathbf{Y}^*)'\mathbf{Y}^* &= \left(\frac{1}{\lambda_1}\text{diag}(\sqrt{\phi_1}, \ldots, \sqrt{\phi_n})\right)' \left(\frac{1}{\lambda_1}\text{diag}(\sqrt{\phi_1}, \ldots, \sqrt{\phi_n})\right) \\ &+ \left(\text{diag}(\sqrt{\phi_1}, \ldots, \sqrt{\phi_n})\right)' \left(\text{diag}(\sqrt{\phi_1}, \ldots, \sqrt{\phi_n})\right) = \frac{\lambda_1^2+1}{\lambda_1^2}\text{diag}(\phi_1, \phi_2, \ldots, \phi_n)\end{aligned}$$

Computation of $\mathbf{\Psi}\mathbf{\Omega}^{-1}\mathbf{\Psi}'$

$$\mathbf{\Psi}\mathbf{\Omega}^{-1}\mathbf{\Psi}' = \begin{bmatrix}\mathbf{I}_{n\times n} & \mathbf{0} & \mathbf{0}\end{bmatrix} \bar{\mathbf{\Omega}}^{-1} \begin{bmatrix}\mathbf{I}_{n\times n} \\ \mathbf{0} \\ \mathbf{0}\end{bmatrix} = \mathbf{I}_{n\times n}\frac{1}{\lambda_1^2}\text{diag}(\phi_1, \ldots, \phi_n)\mathbf{I}_{n\times n} = \frac{1}{\lambda_1^2}\text{diag}(\phi_1, \ldots, \phi_n)$$

Hence,

$$\mathbf{\Phi} = \frac{\lambda_1^2+1}{\lambda_1^2}\text{diag}(\phi_1, \phi_2, \ldots, \phi_n) - \frac{1}{\lambda_1^2}\text{diag}(\phi_1, \ldots, \phi_n) = \frac{\lambda_1^2}{\lambda_1^2}\text{diag}(\phi_1, \ldots, \phi_n) = \text{diag}(\phi_1, \ldots, \phi_n)$$

Let

$$\lambda_1 = \tilde{\lambda}_1,\ \lambda_2 = \tilde{\lambda}_3/2 \text{ and } \lambda_3 = 1$$

where

$$\tilde{\lambda}_1 \text{ and } \tilde{\lambda}_3$$

are the parameters in the Minnesota Prior, then the two approaches give them same result.

Illustration with 2 variables and 2 lags

Consider a VAR(2) model with 2 variables:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} b_{11}^{(1)} & b_{12}^{(1)} \\ b_{21}^{(1)} & b_{22}^{(1)} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{11}^{(2)} & b_{12}^{(2)} \\ b_{21}^{(2)} & b_{22}^{(2)} \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \end{bmatrix} + \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

The system has $m = np + 1 = 2 \times 2 + 1 = 5$ regressors. The dummy observation matrices $\mathbf{Y}^*$ and $\mathbf{X}^*$ are constructed as follows:

Prior for $\mathbf{B}_1$ (First Lag): The first two rows encode the random walk prior:

$$\mathbf{Y}^*[1:2,:] = \begin{bmatrix} \frac{\sqrt{\psi_1}}{\lambda_1} & 0 \\ 0 & \frac{\sqrt{\psi_2}}{\lambda_1} \end{bmatrix}$$

$$\mathbf{X}^*[1:2,:] = \begin{bmatrix} \frac{\sqrt{\psi_1}}{\lambda_1} & 0 & 0 & 0 & 0 \\ 0 & \frac{\sqrt{\psi_2}}{\lambda_1} & 0 & 0 & 0 \end{bmatrix}$$

Prior for $\mathbf{B}_2$ (Second Lag): Rows 3-4 implement the shrinkage prior for the second lag:

$$\mathbf{Y}^*[3:4,:] = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\mathbf{X}^*[3:4,:] = \begin{bmatrix} 0 & 0 & \frac{\sqrt{\psi_1}}{\lambda_1} 2^{\lambda_2} & 0 & 0 \\ 0 & 0 & 0 & \frac{\sqrt{\psi_2}}{\lambda_1} 2^{\lambda_2} & 0 \end{bmatrix}$$

Prior for Constants: Row 5 encodes the prior for the constant terms:

$$\mathbf{Y}^*[5,:] = \begin{bmatrix} 0 & 0 \end{bmatrix}$$

$$\mathbf{X}^*[5,:] = \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{1}{\sqrt{V_c}} \end{bmatrix}$$

Prior for Covariance Matrix: The final $\lambda_3 \times 2$ dummy observations (rows 6 to $5 + 2\lambda_3$) enforce a proper posterior for $\mathbf{\Sigma}$:

$$\mathbf{Y}^*[6:5+2\lambda_3,:] = \begin{bmatrix} \sqrt{\psi_1} & 0 \\ 0 & \sqrt{\psi_2} \\ \sqrt{\psi_1} & 0 \\ 0 & \sqrt{\psi_2} \\ \vdots & \vdots \\ \sqrt{\psi_1} & 0 \\ 0 & \sqrt{\psi_2} \end{bmatrix} \quad (\lambda_3 \text{ repetitions})$$

$$\mathbf{X}^*[6:5+2\lambda_3,:] = \mathbf{0}_{2\lambda_3 \times 5}$$

The final dimensions are: $\mathbf{Y}^*$ is $(5 + 2\lambda_3) \times 2$ and $\mathbf{X}^*$ is $(5 + 2\lambda_3) \times 5$.

References

Doan, T., R. Litterman, and C. Sims (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric reviews 3*(1), 1–100.