Bayesian VARs

Constantino Hevia
UTDT

July 11, 2024

Preliminaries: multivariate densities

- **Multivariate normal** distribution for the random vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_k)'$

$$f(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma) = \frac{|\Sigma|^{-1/2}}{(2\pi)^{k/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right)$$

  where $\boldsymbol{\mu} \in R^k$ is the mean and Σ is the $(k \times k)$ covariance matrix. We write $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \Sigma)$.

Preliminaries: multivariate densities

- **Wishart distribution:** Let $x_i \in R^l$ be i.i.d. with $x_i \sim N(0, \Sigma)$ for $i = 1, 2, \ldots, m$. Let $Z = \sum_{i=1}^m x_i x_i'$ be a random matrix with the sum of the squared terms. Z is distributed as a Wishart with scale parameter Σ and m degrees of freedom with density

$$f(Z|\Sigma, m) = \frac{|Z|^{(m-l-1)/2}}{2^{ml/2} |\Sigma|^{m/2} \Gamma_l(m/2)} \exp\left[-\frac{1}{2}\mathrm{tr}(\Sigma^{-1} Z)\right] \tag{1}$$

  where $\Gamma_l(t) = \pi^{l(l-1)/4} \prod_{i=1}^l \Gamma(t - (i-1)/2)$ is the multivariate Gamma function. We write $Z \sim W_l(\Sigma, m)$

Preliminaries: multivariate densities

- **Inverse Wishart distribution:** We say that Z is distributed as an **inverse Wishart**, denoted by $Z \sim iW(\Psi, m)$, if $Z^{-1} \sim W(\Psi^{-1}, m)$. The density function of the inverse Wishart is
$$f(Z|\Psi, m) = \frac{|\Psi|^{m/2}\,|Z|^{(m+l+1)/2}}{2^{ml/2}\,|\Sigma|^{m/2}\,\Gamma_l(m/2)} \exp\left[-\frac{1}{2}\mathrm{tr}(\Psi Z^{-1})\right]. \tag{2}$$

Preliminaries: properties of the Kronecker product

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$$

$$(AC) \otimes B = (AC) \otimes (BI) = (A \otimes B)(C \otimes I)$$

$$(A + B) \otimes C = A \otimes C + B \otimes C$$

$$(A \otimes B)' = A' \otimes B'$$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$$

$$\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\mathrm{tr}(B)$$

As for the determinant of a Kronecker product, let $A_{n \times n}$ and $B_{m \times m}$, then

$$\det(A \otimes B) = \det(A)^m \det(B)^n$$

Vector Autoregression

- A vector autoregression (**VAR**) is a multivariate autoregressive process of the form

$$Y_t = c + D_1 Y_{t-1} + D_2 Y_{t-2} + \dots + D_p Y_{t-p} + v_t; \quad v_t \sim (0, \Omega)$$

  where

  - Y_t is an $n \times 1$ vector of observed variables,
  - c is an $n \times 1$ vector of constants,
  - D_j are $n \times n$ matrices,
  - Ω is an $n \times n$ covariance matrix.

Alternative representations of a VAR(p)

Two representations of the VAR(p):

- Companion form. Useful for computing moments, forecasting, impulse responses, etc.
- Simultaneous equation form. Useful for evaluating the likelihood function, computing restricted estimates and performing Bayesian estimation.

Companion Form

- Transform the n-variable VAR(p) into an np-variable VAR(1)

Example: consider the VAR(3)

$$Y_t = c + D_1 Y_{t-1} + D_2 Y_{t-2} + D_3 Y_{t-3} + v_t$$

Let

$$\boldsymbol{Y}_t = \begin{bmatrix} Y_t \\ Y_{t-1} \\ Y_{t-2} \end{bmatrix}; \boldsymbol{c} = \begin{bmatrix} c \\ \mathbf{0}_{n\times 1} \\ \mathbf{0}_{n\times 1} \end{bmatrix}; \boldsymbol{D} = \begin{bmatrix} D_1 & D_2 & D_3 \\ \boldsymbol{I}_n & \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & \boldsymbol{I}_n & \mathbf{0}_{n\times n} \end{bmatrix};$$

$$\boldsymbol{v}_t = \begin{bmatrix} v_t \\ \mathbf{0}_{n\times 1} \\ \mathbf{0}_{n\times 1} \end{bmatrix}; \boldsymbol{\Omega} = \begin{bmatrix} \Omega & \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} \end{bmatrix}.$$

Then the VAR(3) can be written as

$$\boldsymbol{Y_t} = \boldsymbol{c} + \boldsymbol{D}\boldsymbol{Y}_{t-1} + \boldsymbol{v}_t$$

Companion Form

- In general, for a VAR(p)

$$Y_t = c + D_1 Y_{t-1} + D_2 Y_{t-2} + \ldots + D_p Y_{t-p} + v_t$$

- Let

$$\underset{np\times 1}{\boldsymbol{Y}_t} = \begin{bmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-p+1} \end{bmatrix}; \quad \underset{np\times 1}{\boldsymbol{c}} = \begin{bmatrix} c \\ \mathbf{0}_{n\times 1} \\ \vdots \\ \mathbf{0}_{n\times 1} \end{bmatrix}; \quad \underset{np\times np}{\boldsymbol{D}} = \begin{bmatrix} D_1 & D_2 & \cdots & D_{p-1} & D_p \\ \boldsymbol{I}_n & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_n & \ddots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \ddots & \mathbf{0} & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{I}_n & \mathbf{0} \end{bmatrix}; \quad \underset{np\times 1}{\boldsymbol{v}_t} = \begin{bmatrix} v_t \\ \mathbf{0}_{n\times 1} \\ \vdots \\ \mathbf{0}_{n\times 1} \end{bmatrix};$$

- Then we can write the VAR(p) as a larger VAR(1)

$$\boldsymbol{Y}_t = \boldsymbol{c} + \boldsymbol{D}\boldsymbol{Y}_{t-1} + \boldsymbol{v}_t \tag{3}$$

Simultaneous equation form

$$Y_t = c + D_1 Y_{t-1} + D_2 Y_{t-2} + \ldots + D_p Y_{t-p} + v_t$$

Transpose the VAR and write as

$$\begin{aligned}
Y_t' =& Y_{t-1}' D_1' + Y_{t-2}' D_2' + \cdots + Y_{t-p}' D_p' + c' + v_t' \\
=& \begin{bmatrix} Y_{t-1}' & Y_{t-2}' & \cdots & Y_{t-p}' & 1 \end{bmatrix} \begin{bmatrix} D_1' \\ D_2' \\ \vdots \\ D_p' \\ c' \end{bmatrix} + v_t' \\
=& X_t' \beta + v_t'
\end{aligned}$$

where $\underset{1\times(np+1)}{X_t'} = \begin{bmatrix} Y_{t-1}' & Y_{t-2}' & \cdots & Y_{t-p}' & 1 \end{bmatrix}$ and $\underset{(np+1)\times n}{\beta} = \begin{bmatrix} D_1' \\ D_2' \\ \vdots \\ D_p' \\ c' \end{bmatrix}$.

Simultaneous equation form

- Note that
$$Y_t' = X_t'\beta + v_t'$$
is a system of n equations, one for each period $t = 1, \ldots, T$.

- We now stack vertically the variables across all the observations. Define
$$\underset{T\times n}{\boldsymbol{Y}} = \begin{bmatrix} Y_1' \\ Y_2' \\ \vdots \\ Y_T' \end{bmatrix}; \underset{T\times(np+1)}{\boldsymbol{X}} = \begin{bmatrix} X_1' \\ X_2' \\ \vdots \\ X_T' \end{bmatrix}; \text{and } \underset{T\times n}{\boldsymbol{V}} = \begin{bmatrix} v_1' \\ v_2' \\ \vdots \\ v_T' \end{bmatrix};$$

- We can write the system as
$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{V}. \tag{4}$$

Simultaneous equation form

- Vectorize (i.e. stack the columns) the expression

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{V}$$

and use $\text{vec}\,(ABC) = (C' \otimes A)\,\text{vec}(B)$ to write

$$\text{vec}(\boldsymbol{Y}) = (I_n \otimes \boldsymbol{X})\,\text{vec}(\boldsymbol{\beta}) + \text{vec}(\boldsymbol{V})$$

or

$$y = (I_n \otimes \boldsymbol{X})\,\beta + \nu \tag{5}$$

where $\underset{nT\times 1}{y} = vec(\boldsymbol{Y})$, $\underset{(np+1)n\times 1}{\beta} = vec(\boldsymbol{\beta})$, and $\underset{nT\times 1}{v} = vec(\boldsymbol{V}) \sim N\,(0, \Omega \otimes I_T)$.

Exercise 3: Prove that $E\,[vv'] = \Omega \otimes I_T$.

Simultaneous equation form

- We will write the likelihood function in 3 equivalent ways.

- The likelihood function is

$$L(\beta, \Omega | Y) = \frac{|\ \Omega \otimes I_T\ |^{-0.5}}{(2\pi)^{0.5nT}} \exp\left[-\frac{1}{2}(y - (I_n \otimes \boldsymbol{X})\beta)'(\Omega \otimes I_T)^{-1}(y - (I_n \otimes \boldsymbol{X})\beta)\right]$$

- **Facts:** Let $A_{n\times n}$ and $B_{m\times m}$. Then $|A \otimes B| = |A|^m\,|B|^n$ and $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.

- **Form 1:** Using the previous facts, we can write the likelihood function as

$$L(\beta, \Omega | Y) = \frac{|\Omega|^{-T/2}}{(2\pi)^{nT/2}} \exp\left[-\frac{1}{2}(y - (I_n \otimes \boldsymbol{X})\beta)'(\Omega^{-1} \otimes I_T)(y - (I_n \otimes \boldsymbol{X})\beta)\right]. \quad (6)$$

Simultaneous equation form

- **Facts:** For comformable matrices B, C, and D, we have

$$
\begin{aligned}
\mathrm{vec}(BCD) &= \left(D' \otimes B\right) \mathrm{vec}(C) \\
\mathrm{tr}(B'C) &= \mathrm{vec}(B)'\mathrm{vec}(C) \\
\mathrm{tr}(BCD) &= \mathrm{tr}(CDB) = \mathrm{tr}(DBC).
\end{aligned}
$$

- Using that $y - (I_n \otimes \boldsymbol{X})\beta = \mathrm{vec}(\boldsymbol{V})$, we can write the quadratic form in (6) as

$$
\begin{aligned}
(y - (I_n \otimes \boldsymbol{X})\beta)'(\Omega^{-1} \otimes I_T)(y - (I_n \otimes \boldsymbol{X})\beta) &= \mathrm{vec}(\boldsymbol{V})'(\Omega^{-1} \otimes I_T)\mathrm{vec}(\boldsymbol{V}) \\
&= \mathrm{vec}(\boldsymbol{V})'\mathrm{vec}(I_T \boldsymbol{V}\Omega^{-1}) \\
&= \mathrm{vec}(\boldsymbol{V})'\mathrm{vec}(\boldsymbol{V}\Omega^{-1}) \\
&= \mathrm{tr}(\boldsymbol{V}'\boldsymbol{V}\Omega^{-1}) \\
&= \mathrm{tr}(\Omega^{-1}\boldsymbol{V}'\boldsymbol{V}).
\end{aligned}
$$

Simultaneous equation form

- **Form 2**: then we can write the likelihood function as

$$L(\beta,\Omega|Y)=\frac{|\,\Omega\,|^{-T/2}}{(2\pi)^{nT/2}}\exp\left[-\frac{1}{2}\mathrm{tr}(\Omega^{-1}\boldsymbol{V}'\boldsymbol{V})\right]. \tag{7}$$

- **Exercise 4**: Consider the term in the exponent in (6). Show that

$$(y-(I_n\otimes\boldsymbol{X})\beta)'(\Omega^{-1}\otimes I_T)(y-(I_n\otimes\boldsymbol{X})\beta)=$$
$$y'(\Omega^{-1}\otimes I_T)y-2\beta'(\Omega^{-1}\otimes\boldsymbol{X}')y+\beta'(\Omega^{-1}\otimes\boldsymbol{X}'\boldsymbol{X})\beta.$$

- **Exercise 5**: Using the previous result, show that the MLE estimator of β is equal to the OLS estimator of β and given by

$$\beta^{\text{mle}}=\left(\Omega\otimes(\boldsymbol{X}'\boldsymbol{X})^{-1}\right)\left(\Omega^{-1}\otimes\boldsymbol{X}'\right)y=\left(I_n\otimes(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\right)y=\beta^{\text{ols}}. \tag{8}$$

Yet another way of writing the likelihood

Some manipulation of the quadratic term in (6)

$$
\begin{aligned}
(y-(I_n\otimes \boldsymbol{X})\beta)'(\Omega^{-1}\otimes I_T)(y-(I_n\otimes \boldsymbol{X})\beta) &= \\
(y-(I_n\otimes \boldsymbol{X})\beta)'(\Omega^{-0.5}\otimes I_T)(\Omega^{-0.5}\otimes I_T)(y-(I_n\otimes \boldsymbol{X})\beta) &= \\
\left[(\Omega^{-0.5}\otimes I_T)(y-(I_n\otimes \boldsymbol{X})\beta)\right]'\left[(\Omega^{-0.5}\otimes I_T)(y-(I_n\otimes \boldsymbol{X})\beta)\right] &= \\
\left[(\Omega^{-0.5}\otimes I_T)y-(\Omega^{-0.5}\otimes I_T)(I_n\otimes \boldsymbol{X})\beta\right]'\left[(\Omega^{-0.5}\otimes I_T)y-(\Omega^{-0.5}\otimes I_T)(I_n\otimes \boldsymbol{X})\beta\right] &= \\
\left[(\Omega^{-0.5}\otimes I_T)y-(\Omega^{-0.5}\otimes \boldsymbol{X})\beta\right]'\left[(\Omega^{-0.5}\otimes I_T)y-(\Omega^{-0.5}\otimes \boldsymbol{X})\beta\right]&.
\end{aligned}
$$

where in the last line we used

$$
(\Omega^{-0.5}\otimes I_T)(I_n\otimes \boldsymbol{X}) = \Omega^{-0.5}\otimes \boldsymbol{X}
$$

Yet another way of writing the likelihood

- Now consider the term

$$
\begin{aligned}
(\Omega^{-0.5} \otimes I_T)y - (\Omega^{-0.5} \otimes \boldsymbol{X})\beta &= \\
(\Omega^{-0.5} \otimes I_T)y - (\Omega^{-0.5} \otimes \boldsymbol{X})\beta + (\Omega^{-0.5} \otimes \boldsymbol{X})\beta^{\text{ols}} - (\Omega^{-0.5} \otimes \boldsymbol{X})\beta^{\text{ols}} &= \\
\underbrace{(\Omega^{-0.5} \otimes I_T)y - (\Omega^{-0.5} \otimes \boldsymbol{X})\beta^{\text{ols}}}_{=W} + \underbrace{(\Omega^{-0.5} \otimes \boldsymbol{X})\left(\beta^{\text{ols}} - \beta\right)}_{\equiv Z} &= W + Z.
\end{aligned}
$$

- It then follows that the quadratic term is

$$
\begin{aligned}
(y - (I_n \otimes \boldsymbol{X})\beta)'(\Omega^{-1} \otimes I_T)(y - (I_n \otimes \boldsymbol{X})\beta) &= (W + Z)'(W + Z) \\
&= W'W + 2Z'W + Z'Z.
\end{aligned}
$$

Yet another way of writing the likelihood

- **Exercise 6**: Show that

$$
\begin{aligned}
W'W =& \operatorname{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\\
Z'Z =& \left(\beta-\beta^{\text{ols}}\right)'\left(\Omega^{-1}\otimes(\boldsymbol{X}'\boldsymbol{X})\right)\left(\beta-\beta^{\text{ols}}\right)\\
Z'W =& 0.
\end{aligned}
$$

  where $\boldsymbol{S}^{\text{ols}} = (\boldsymbol{Y}-\boldsymbol{X}\beta^{\text{ols}})'(\boldsymbol{Y}-\boldsymbol{X}\beta^{\text{ols}})$ is the sum of squared residuals from the OLS estimation.

- Hence, the quadratic term in the likelihood function (6) can be written as

$$
(y-(I_n\otimes\boldsymbol{X})\beta)'(\Omega^{-1}\otimes I_T)(y-(I_n\otimes\boldsymbol{X})\beta) = \\
\left(\beta-\beta^{\text{ols}}\right)'\left(\Omega^{-1}\otimes\boldsymbol{X}'\boldsymbol{X}\right)\left(\beta-\beta^{\text{ols}}\right)+\operatorname{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)
$$

Yet another way of writing the likelihood

- Therefore, the likelihood function (6) can be written as

$$L(\beta, \Omega|Y) = \frac{|\Omega|^{-T/2}}{(2\pi)^{nT/2}} \exp\left[-\frac{1}{2}\left(\beta - \beta^{\text{ols}}\right)'(\Omega^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})\left(\beta - \beta^{\text{ols}}\right)\right] \times \exp\left[-\frac{1}{2}\text{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\right]$$

- **Form 3**: Doing some algebra, we can write

$$L(\beta, \Omega|Y) \propto \frac{|\Omega \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}|^{-1/2}}{(2\pi)^{kn/2}} \exp\left[-\frac{1}{2}\left(\beta - \beta^{\text{ols}}\right)'(\Omega^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})\left(\beta - \beta^{\text{ols}}\right)\right] \\ \times |\Omega|^{-\frac{(T-k-n-1)+n+1}{2}} \exp\left[-\frac{1}{2}\text{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\right] \quad (9)$$

- The term in blue is the density of a multivariate normal for $\beta \sim N(\beta^{\text{ols}}, \Omega \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1})$. The term in red is the kernel of an inverse Wishart for $\Omega \sim iW(\boldsymbol{S}^{\text{ols}}, T-k-n-1)$.

Tree equivalent ways of writing the likelihood

The likelihood function can be written in any of the following forms

$$L(\beta, \Omega | Y) = \frac{|\Omega|^{-T/2}}{(2\pi)^{nT/2}} \exp\left[-\frac{1}{2}(y - (I_n \otimes \boldsymbol{X})\beta)'(\Omega^{-1} \otimes I_T)(y - (I_n \otimes \boldsymbol{X})\beta)\right]$$

$$= \frac{|\ \Omega\ |^{-T/2}}{(2\pi)^{nT/2}} \exp\left[-\frac{1}{2}\text{tr}(\Omega^{-1}\boldsymbol{V}'\boldsymbol{V})\right]$$

$$\propto \frac{\left|\Omega \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\right|^{-1/2}}{(2\pi)^{kn/2}} \exp\left[-\frac{1}{2}\left(\beta - \beta^{\text{ols}}\right)'(\Omega^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})\left(\beta - \beta^{\text{ols}}\right)\right]$$
$$\times |\Omega|^{-\frac{(T-k-n-1)+n+1}{2}} \exp\left[-\frac{1}{2}\text{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\right].$$

Priors for vector autoregressions

The following are common priors for VARs

1. Normal prior for β assuming that Ω is known (Theil mixed estimator).
2. Non-informative (diffuse) prior for both β and Ω (Jeffreys prior).
3. Normal prior for β and non-informative prior for Ω.
4. Normal prior for β and independent inverse Wishart prior for Ω.

Illustration: VAR for US inflation, unemployoyent, and interest rate

- To analyze the four priors discussed above, we will consider a VAR for US inflation, unemployment, and the interest rate.

- Quarterly model with $Y_t = (\text{inflation}_t, \text{unemployment}_t, \text{interest rate}_t)$.

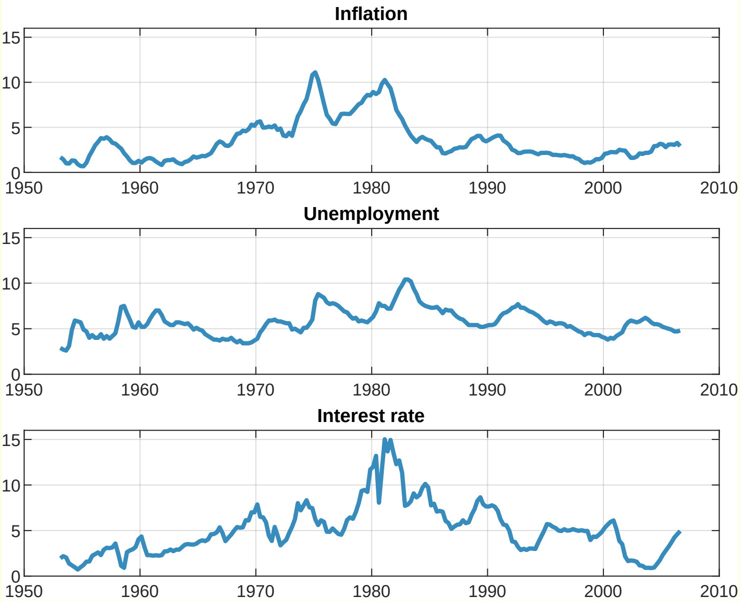
- Data span: 1953:Q1 through 2006:Q3.

- Data modeled as a VAR(4)

$$Y_t = c + D_1 Y_{t-1} + D_2 Y_{t-2} + D_3 Y_{t-3} + D_4 Y_{t-4} + v_t$$

- Bayesian impulse responses to an interest rate shock identified using Cholesky.

Illustration: US inflation, unemployent, and interest rate



Case 1: Normal prior for β with known Ω (Theil mixed estimator)

- Assume that Ω is known. To implement the procedure, we set $\Omega = \Omega^{ols}$ as an approximation.

- Assume the prior for $\beta \sim N(\beta_0, V_0)$, where β_0 is $nk \times 1$ and V_0 is $nk \times nk$:

$$p(\beta) \propto \exp\left(-\frac{1}{2}(\beta - \beta_0)' V_0^{-1} (\beta - \beta_0)\right). \tag{10}$$

- Use **Form 1** of the likelihood function (6) with Ω known:

$$L(\beta|\Omega, Y) = \frac{|\Omega|^{-T/2}}{(2\pi)^{nT/2}} \exp\left[-\frac{1}{2}(y - (I_n \otimes \boldsymbol{X})\beta)'(\Omega^{-1} \otimes I_T)(y - (I_n \otimes \boldsymbol{X})\beta)\right]$$

Case 1: Normal prior for β with known Ω (Theil mixed estimator)

- The posterior satisfies

$$
\begin{aligned}
p(\beta|\Omega, Y) \propto & L(\beta|\Omega, Y)p(\beta) \\
& \propto \exp\left[-\frac{1}{2}(y-(I_n\otimes \boldsymbol{X})\beta)'(\Omega^{-1}\otimes I_T)(y-(I_n\otimes \boldsymbol{X})\beta)\right] \\
& \quad \times \exp\left[-\frac{1}{2}(\beta-\beta_0)'V_0^{-1}(\beta-\beta_0)\right] \\
& \propto \exp\left[\frac{1}{2}\left((y-(I_n\otimes \boldsymbol{X})\beta)'(\Omega^{-1}\otimes I_T)(y-(I_n\otimes \boldsymbol{X})\beta)+(\beta-\beta_0)'V_0^{-1}(\beta-\beta_0)\right)\right]
\end{aligned}
$$

- The exponent has a sum of two quadratic terms in β. Therefore, it is also a quadratic form in β:

$$
(\beta-\beta_1)'V_1^{-1}(\beta-\beta_1) = \beta' V_1^{-1}\beta - 2\beta_1' V_1^{-1}\beta + \beta_1' V_0^{-1}\beta_1 \tag{11}
$$

- We want to find the parameters β_1 and V_1 of the new quadratic form. This methodology is called **completing the square** and will be used to find the posterior distributions of β.

Case 1: Normal prior for β with known Ω (Theil mixed estimator)

- First note that

$$(\beta-\beta_0)'V_0^{-1}(\beta-\beta_0)=\beta'V_0^{-1}\beta-2\beta_0'V_0^{-1}\beta+\beta_0'V_0^{-1}\beta_0.$$

- Likewise

$$(y-(I_n\otimes \boldsymbol{X})\beta)'(\Omega^{-1}\otimes I_T)(y-(I_n\otimes \boldsymbol{X})\beta)=y'(\Omega^{-1}\otimes I_T)y-2y'(\Omega^{-1}\otimes \boldsymbol{X})\beta+\beta'(\Omega^{-1}\otimes(\boldsymbol{X}'\boldsymbol{X}))\beta$$

- Summing both terms we have

$$\begin{aligned}(y-(I_n\otimes \boldsymbol{X})\beta)'(\Omega^{-1}\otimes I_T)(y-(I_n\otimes \boldsymbol{X})\beta)+(\beta-\beta_0)'V_0^{-1}(\beta-\beta_0)&=\\ \beta'\left[V_0^{-1}+\Omega^{-1}\otimes(\boldsymbol{X}'\boldsymbol{X})\right]\beta-2\left[\beta_0'V_0^{-1}+y'(\Omega^{-1}\otimes \boldsymbol{X})\right]\beta+\beta_0'V_0^{-1}\beta_0+y'(\Omega^{-1}\otimes I_T)y&=\\ \beta'\left[V_0^{-1}+\Omega^{-1}\otimes(\boldsymbol{X}'\boldsymbol{X})\right]\beta-2\left[\beta_0'V_0^{-1}+y'(\Omega^{-1}\otimes \boldsymbol{X})\right]\beta+C&.\end{aligned}$$

- We can ignore the green term because it does not contain the random variable β.

Case 1: Normal prior for β with known Ω (Theil mixed estimator)

- Therefore, we need to compare

$$\beta' \left[V_0^{-1} + \Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X}) \right] \beta - 2 \left[\beta_0' V_0^{-1} + y'(\Omega^{-1} \otimes \boldsymbol{X}) \right] \beta$$

  with the quadratic form that we are looking for, which has the form

$$\beta' V_1^{-1} \beta - 2\beta_1' V_1^{-1} \beta + \beta_1' V_0^{-1} \beta_1$$

- Matching coefficients, we need

$$\begin{aligned} V_1^{-1} &= \left[V_0^{-1} + \Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X}) \right] \\ \beta_1' V_1^{-1} &= \left[\beta_0' V_0^{-1} + y'(\Omega^{-1} \otimes \boldsymbol{X}) \right]. \end{aligned}$$

- We can rewrite these equations as

$$\begin{aligned} V_1 &= \left[V_0^{-1} + \Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X}) \right]^{-1} \\ \beta_1 &= V_1 \left[V_0^{-1} \beta_0 + (\Omega^{-1} \otimes \boldsymbol{X}') y \right]. \end{aligned}$$

Case 1: Normal prior for β with known Ω (Theil mixed estimator)

- **Exercise 6:** Use equation (8) to prove that we can write

$$\beta_1 = \left[V_0^{-1} + \Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right]^{-1} \left[V_0^{-1}\beta_0 + (\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X}))\beta^{\text{ols}}\right] = W\beta_0 + (I - W)\beta^{\text{ols}}$$

  where $W = \left[V_0^{-1} + \Omega^{-1} \otimes \boldsymbol{X}'\boldsymbol{X}\right]^{-1} V_0^{-1}$. That is, β_1 is a weighted average of the prior β_0 and the OLS estimate β^{ols}.

- Therefore, the posterior distribution is proportional to

$$p(\beta|\Omega, Y) \propto \exp\left[(\beta - \beta_1)' V_1^{-1} (\beta - \beta_1)\right]$$

  which is the kernel of a multivariate normal distribution with mean and covariance

$$V_1 = \left[V_0^{-1} + \Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X})\right]^{-1} \tag{12}$$

$$\beta_1 = V_1 \left[V_0^{-1}\beta_0 + (\Omega^{-1} \otimes (\boldsymbol{X}'\boldsymbol{X}))\beta^{\text{ols}}\right]. \tag{13}$$

Case 1: Normal prior for β with known Ω (Theil mixed estimator)

- Summarizing, given the parameters of the prior distribution β_0 and V_0, the data $\boldsymbol{X}$ and y, and the known covariance matrix Ω, we use equations (12) and (13) to update the covariance and the mean of the posterior distribution, V_1 and β_1.

- To implement this procedure, we do the following:

  1. Choose the parameters of the prior distribution of β: β_0 and V_0.
  2. Set $\Omega = \Omega^{\text{ols}}$, the OLS estimate of the covariance matrix.
  3. Draw from the posterior distribution of β from a normal distribution $N(\beta_1, V_1)$, where β_1 and V_1 satisfy equations (12) and (13).
  4. Repeat many times and compute whatever you want, like impulse response functions, forecasts, etc.

- This procedure requires us to choose a large number of parameters: nk in β_0 and $nk(nk-1)/2$ in V_0. This is too much.

A special case of Case 1: the Minnesota prior

- In the Minnesota (or Litterman) prior, β_0 and V_0 are functions of a small number of hyperparameters.

- The Minnesota prior has been useful in forecasting persistent economic time series.

- Traditional implementation:
  1. Set the prior mean of the first lag of the dependent variable to one and all the other prior means to zero. Prior is that each variable follows an independent random walk:
  $$y_{i,t} = y_{i,t-1} + v_{i,t}.$$
  2. Set the prior variance of the ij element of the matrix D_ℓ as
  $$v^D_{ij,\ell} = \begin{cases} \frac{\lambda_1}{\ell^{\lambda_3}} & \text{if } i = j \\ \frac{\lambda_1 \lambda_2}{\ell^{\lambda_3}} \left(\frac{\sigma_i}{\sigma_j}\right)^2 & \text{if } i \neq j. \end{cases}$$
  3. Set the prior variances of exogenous variables (i.e. intercepts in c) to :
  $$v^D_i = \lambda_1 \lambda_4$$

A special case of Case 1: the Minnesota prior

- λ_1 represents the tightness on the variance of the first lag on own variable. It is the prior variance of $D_1[i, i]$.

- λ_2 controls the relative tightness of other variables $j \neq i$ in equation i. Usually, $0 < \lambda_2 \leq 1$.

- ℓ^{λ_3} controls the tightness of the variance of lags ℓ other than the first one.

- σ_i is the i^{th} diagonal element of Ω. It may also set as the variance of the error term on univariate autoregressions on the variable y_i.

- λ_4 is the relative tightness of the exogenous variables (i.e. intercept). Usually λ_4 is set to a large number.

A special case of Case 1: the Minnesota prior

- **Rule of thumb:** $\lambda_1 = 0.2$, $\lambda_2 = 0.5$, $\lambda_3 = 1$ or 2, $\lambda_4 = 10^5$.

- **Logic of the Minnesota prior**: The *n* time series are *a priori* represented as a random walk because univariate random walks are typically good at forecasting persistent macroeconomic series.

- **Important:** if the variables are not very persistent (for example, growth rates), assuming a random walk may not be the best choice. In this case the prior may be that all coefficients on lagged values in β_0 are zero.

Example: bivariate VAR(2) model

- Consider the following bivariate VAR(2)

$$
\begin{aligned}
y_{1,t} &= \underset{(\lambda_1\lambda_4)}{0} + \underset{(\lambda_1)}{1 \cdot y_{1,t-1}} + \underset{(\lambda_1\lambda_2\sigma_1^2/\sigma_2^2)}{0 \cdot y_{2,t-1}} + \underset{(\lambda_1/2^{\lambda_3})}{0 \cdot y_{1,t-2}} + \underset{(\lambda_1\lambda_2\sigma_1^2/(2^{\lambda_3}\sigma_2^2))}{0 \cdot y_{2,t-2}} + v_{1,t} \\
y_{2,t} &= \underset{(\lambda_1\lambda_4)}{0} + \underset{(\lambda_1\lambda_2\sigma_2^2/\sigma_1^2)}{0 \cdot y_{1,t-1}} + \underset{(\lambda_1)}{1 \cdot y_{2,t-1}} + \underset{(\lambda_1\lambda_2\sigma_2^2/(2^{\lambda_3}\sigma_1^2))}{0 \cdot y_{1,t-2}} + \underset{(\lambda_1/2^{\lambda_3})}{0 \cdot y_{2,t-2}} + v_{2,t}.
\end{aligned}
$$

Example: bivariate VAR(2) model

The prior mean is thus

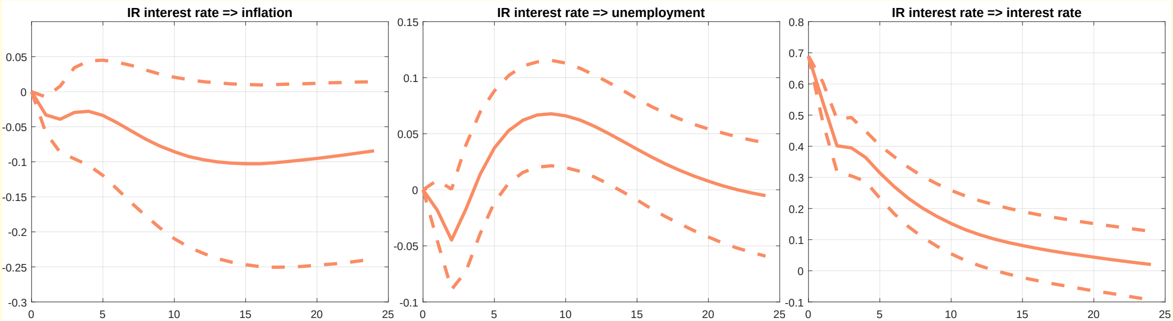
$$\beta_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow \begin{bmatrix} D_1[1,1] \\ D_1[1,2] \\ D_2[1,1] \\ D_2[1,1] \\ c_1 \\ D_1[2,1] \\ D_1[2,2] \\ D_2[2,1] \\ D_2[2,2] \\ c_2 \end{bmatrix} = \beta$$

Example: bivariate VAR(2) model

The prior variance is

$$V_0 = \begin{bmatrix} \lambda_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_1\lambda_2\frac{\sigma_1^2}{\sigma_2^2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\lambda_1}{2^{\lambda_3}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\lambda_1\lambda_2}{2^{\lambda_3}}\frac{\sigma_1^2}{\sigma_2^2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_1\lambda_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_1\lambda_2\frac{\sigma_2^2}{\sigma_1^2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\lambda_1\lambda_2}{2^{\lambda_3}}\frac{\sigma_2^2}{\sigma_1^2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{\lambda_1}{2^{\lambda_3}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1\lambda_4 \end{bmatrix}$$

Impuse Response to interest rate shock using the Minnesota Prior



Case 2: Diffuse prior for β and Ω

- Jeffreys (1961) proposed a rule for generating priors that are non-informative about the parameter of interest and retain certain useful properties.

- The Jeffreys' prior is proportional to the square root of the determinant of the Fisher information matrix.

- In the case of a VAR with n variables, the diffuse prior is

$$p(\beta, \Omega) = p(\beta)p(\Omega)$$

  where

$$p(\beta) = constant$$
$$p(\Omega) \propto |\Omega|^{-(n+1)/2}.$$

- The posterior distribution is $p(\beta, \Omega|Y) \propto L(\beta, \Omega|Y)p(\beta)p(\Omega)$

Case 2: Diffuse prior for β and Ω

- Using **Form 3** of the likelihood in equation (9) we get

$$
\begin{aligned}
p(\beta, \Omega | Y) \propto & \frac{\left|\Omega \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\right|^{-1/2}}{(2\pi)^{kn/2}} \exp\left[-\frac{1}{2}\left(\beta - \beta^{\text{ols}}\right)' \left(\Omega^{-1} \otimes \boldsymbol{X}'\boldsymbol{X}\right)\left(\beta - \beta^{\text{ols}}\right)\right] \times \\
& |\Omega|^{-\frac{n+1}{2}} |\Omega|^{-\frac{(T-k-n-1)+n+1}{2}} \exp\left[-\frac{1}{2}\text{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\right] \\
& \propto N\left(\beta^{\text{ols}}, \Omega \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\right) |\Omega|^{-\frac{(T-k)+n+1}{2}} \exp\left[-\frac{1}{2}\text{tr}\left(\boldsymbol{S}^{\text{ols}}\Omega^{-1}\right)\right] \\
& \propto N\left(\beta^{\text{ols}}, \Omega \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\right) \times iW(\Omega | \boldsymbol{S}^{\text{ols}}, T-k).
\end{aligned}
$$

- Integrating over β we get

$$
p(\Omega|Y) = \int_\beta p(\beta, \Omega|Y) = iW(\Omega|\boldsymbol{S}^{\text{ols}}, T-k). \tag{14}
$$

Case 2: Diffuse prior for β and Ω

- Therefore, the conditional probability of $\beta| \{\Omega, Y\}$ is

$$p(\beta|\Omega, Y) = p(\beta, \Omega|Y)/p(\Omega|Y) = N\left(\beta^{\text{ols}}, \Omega \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\right)$$

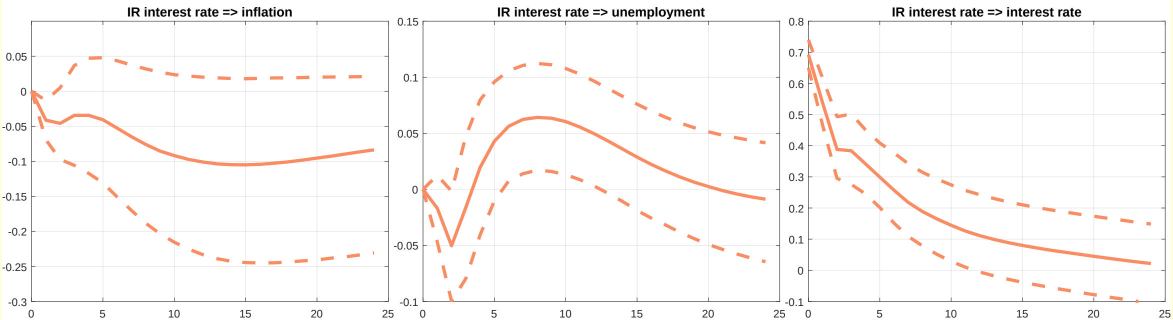
- We computed the conditionals $p(\Omega|Y)$ and $p(\beta|\Omega, Y)$.

- To draw from the posterior $p(\beta, \Omega|Y)$ we can use Gibbs Sampler to construct a Markov Chain $\{\beta^j, \Omega^j\}$ with invariant distribution $p(\beta, \Omega|Y)$.

- **Algorithm:** `Choose` N `large and set` $j = 1$. `Then do the following:`

  1. `Draw` $\Omega^j \sim iW(\Omega|\boldsymbol{S}^{\text{ols}}, T-k)$.
  2. `Draw` $\beta^j \sim N\left(\beta^{\text{ols}}, \Omega^j \otimes (\boldsymbol{X}'\boldsymbol{X})^{-1}\right)$.
  3. `Store` $\{\beta^j, \Omega^j\}$, `set` $j = j+1$ `and return to step 1 while` $j < N$.
  4. `Discard an initial burn in sample and compute whatever you want.`

IR to interest rate shock using Diffuse priors for β and Ω



Case 3: Normal prior for β, diffuse prior for Ω

- Here we assume
$$p(\beta, \Omega) = p(\beta)p(\Omega)$$
where
$$\begin{aligned} p(\beta) &= N(\beta_0, V_0) \\ p(\Omega) &\propto |\Omega|^{-(n+1)/2}. \end{aligned}$$

- To analyze this case, we use the second form of the likelihood (7)
$$\begin{aligned} L(\beta, \Omega|Y) &= \frac{|\Omega|^{-T/2}}{(2\pi)^{nT/2}} \exp\left[-\frac{1}{2}\mathrm{tr}(\Omega^{-1}\boldsymbol{V}'\boldsymbol{V})\right] \\ &\propto |\Omega|^{-T/2} \exp\left[-\frac{1}{2}\mathrm{tr}((\boldsymbol{Y}-\boldsymbol{XB})'(\boldsymbol{Y}-\boldsymbol{XB}))\Omega^{-1}\right] \\ &\propto |\Omega|^{-T/2} \exp\left[-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}\Omega^{-1})\right] \end{aligned}$$
where $\boldsymbol{S} = (\boldsymbol{Y}-\boldsymbol{XB})'(\boldsymbol{Y}-\boldsymbol{XB})$ is sum of residual squares when the parameter is $\boldsymbol{B}$. (Recall $\beta = \mathrm{vec}\boldsymbol{B}$).

Case 3: Normal prior for β, diffuse prior for Ω

- Then, if β is known, the posterior of $p(\Omega|\beta, Y) = L(\beta, \Omega|Y)p(\Omega)$ is

$$
\begin{aligned}
p(\Omega|\beta, Y) \propto & |\Omega|^{-T/2} \exp\left[-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}\Omega^{-1})\right] \times |\Omega|^{-(n+1)/2} \\
\propto & |\Omega|^{-(T+n-1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}\Omega^{-1})\right] \\
= & iW(\boldsymbol{S}, T).
\end{aligned}
$$

- Now suppose that Ω is known. Then following the steps of Case 1, we conclude that

$$p(\beta|\Omega, Y) = N(\beta_1, V_1)$$

where

$$
\begin{aligned}
V_1 = & \left[V_0^{-1} + \Omega^{-1} \otimes \boldsymbol{X}'\boldsymbol{X}\right]^{-1} \\
\beta_1 = & V_1 \left[V_0^{-1}\beta_0 + (\Omega^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})\beta^{\mathrm{ols}}\right].
\end{aligned}
$$

Case 3: Normal prior for β, diffuse prior for Ω

- Since we have the conditional posteriors $p(\Omega|\beta, Y)$ and $p(\beta|\Omega, Y)$, we can draw from the posterior $p(\beta, \Omega|Y)$ using the Gibbs Sampler.

- **Algorithm:** Choose N large, set $j = 1$, and choose an initial $\beta^{(0)}$. Then do the following:

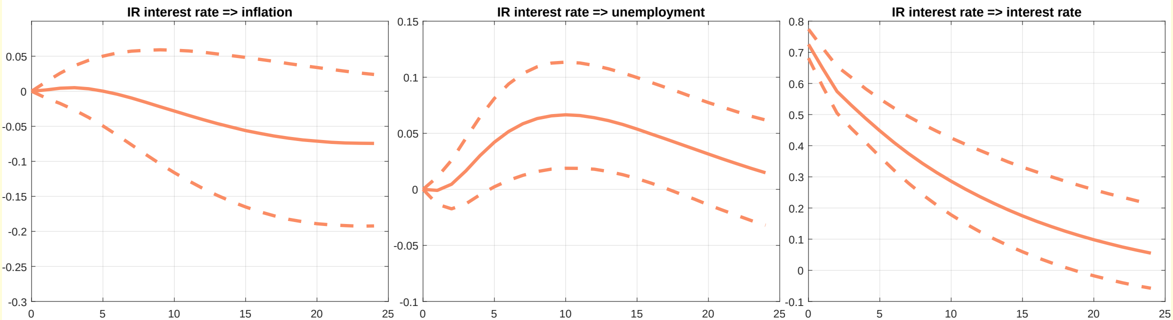
  1. Draw $\Omega^j \sim iW(\Omega|\boldsymbol{S}^{(j-1)}, T)$, where $\boldsymbol{S}^{(j-1)} = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}^{(j-1)})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}^{(j-1)})$ is the sum of squared residuals using $\beta^{(j-1)}$ as the VAR parameters.

  2. Given $\Omega^{(j)}$ from the previous step, draw $\beta^j \sim p(\beta|\Omega^{(j)}, Y)$ using the previous formulas.

  3. Store $\{\beta^j, \Omega^j\}$, set $j = j + 1$ and return to step 1 while $j < N$.

  4. Discard an initial burn in sample and compute whatever you want.

IR to interest rate shock using Normal prior for β and diffuse for Ω



Case 4: Independent Normal-Wishart priors

- Here we assume
$$p(\beta, \Omega) = p(\beta)p(\Omega)$$
where
$$\begin{aligned} p(\beta) =& N(\beta_0, V_0) \propto \exp\left[-\frac{1}{2}(\beta - \beta_0)' V_0^{-1}(\beta - \beta_0)\right] \\ p(\Omega) \propto& iW(S_0, v_0) \propto |\Omega|^{-\frac{v_0+n+1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}_0\Omega^{-1})\right]. \end{aligned}$$

- To compute $p(\Omega|\beta, Y)$, we use, as in Case 3, the second form of the likelihood (7)
$$L(\beta, \Omega|Y) \propto |\Omega|^{-T/2} \exp\left[-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}\Omega^{-1})\right]$$
where $\boldsymbol{S} = (\boldsymbol{Y} - \boldsymbol{XB})'(\boldsymbol{Y} - \boldsymbol{XB})$ is sum of residual squares when the parameter is $\boldsymbol{B}$.

Case 4: Independent Normal-Wishart priors

- Then the conditional posterior for Ω satisfies

$$
\begin{aligned}
p(\Omega|\beta, Y) \propto & |\Omega|^{-T/2} \exp\left[-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}\Omega^{-1})\right] \times |\Omega|^{-\frac{v_0+n+1}{2}} \exp\left[-\frac{1}{2}\mathrm{tr}(\boldsymbol{S}_0\Omega^{-1})\right] \\
\propto & |\Omega|^{-(T+v_0+n-1)/2} \exp\left[-\frac{1}{2}\mathrm{tr}((\boldsymbol{S}+\boldsymbol{S}_0)\Omega^{-1})\right] \\
= & iW(\boldsymbol{S}+\boldsymbol{S}_0, T+v_0).
\end{aligned}
$$

- As in Case 1 and 3, conditional on Ω, Y the posterior of β is

$$p(\beta|\Omega, Y) = N(\beta_1, V_1)$$

where

$$
\begin{aligned}
V_1 = & \left[V_0^{-1} + \Omega^{-1} \otimes \boldsymbol{X}'\boldsymbol{X}\right]^{-1} \\
\beta_1 = & V_1\left[V_0^{-1}\beta_0 + (\Omega^{-1} \otimes \boldsymbol{X}'\boldsymbol{X})\beta^{\text{ols}}\right].
\end{aligned}
$$

Case 3: Normal prior for β, diffuse prior for Ω

- Since we have the conditional posteriors $p(\Omega|\beta, Y)$ and $p(\beta|\Omega, Y)$, we can draw from the posterior $p(\beta, \Omega|Y)$ using the Gibbs Sampler.

- **Algorithm**: Choose N large, set $j = 1$, and choose an initial $\beta^{(0)}$. Then do the following:

  1. Draw $\Omega^j \sim iW(\Omega|(\boldsymbol{S}^{(j-1)} + \boldsymbol{S}_0), T + v_0)$, where $\boldsymbol{S}^{(j-1)} = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}^{(j-1)})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{B}^{(j-1)})$ is the sum of squared residuals using $\beta^{(j-1)}$ as the VAR parameters.

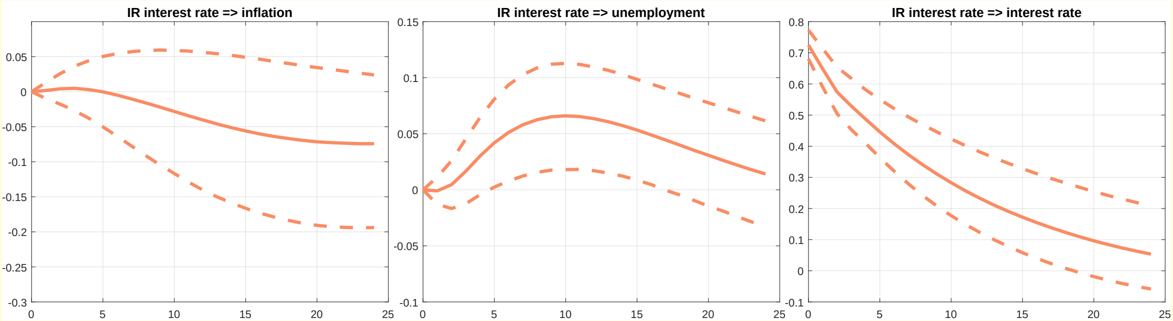
  2. Given $\Omega^{(j)}$ from the previous step, draw $\beta^j \sim p(\beta|\Omega^{(j)}, Y)$ using the previous formulas.

  3. Store $\{\beta^j, \Omega^j\}$, set $j = j + 1$ and return to step 1 while $j < N$.

  4. Discard an initial burn in sample and compute whatever you want.

- Relative to Case 3, only step 1 of the algorithm changes by drawing from a different inverse Wishart.

IR to interest rate shock using the independent Normal-Wishart prior



Software

- Your programming skills.
  - Advantage: you know what you are doing. You are in control of everything.
  - Disadvantages: Time consuming and error prone.
- **BEAR toolbox**:
  - The Bayesian Estimation, Analysis and Regression toolbox (BEAR) is a comprehensive (Bayesian Panel) VAR toolbox for forecasting and policy analysis.
  - Matlab toolbox with a user-friendly interface.
  - Developed and maintained by researchers at the European Central Bank.
  - Does everything we discussed and more (stochastic volatility, TVP-VAR, panel-VAR, etc)
  - Download and find more information here.