Bayesian Econometrics
Tutorial 05 - Bayesian VAR Models

Tutor: Richard Schnorrenberger

richard.schn@stat-econ.uni-kiel.de

Institute for Statistics and Econometrics

Kiel University

Winter Term 2024/25

Bayesian VAR Models

- The Vector Autoregressive (VAR) model is the most popular tool in empirical macroeconomics for policy analysis and forecasting.
- VAR models are regression models with many equations where the dependent variables depend on their own past values and also on lags of all other variables.
- VAR models tend to be overparameterized: sample information alone is not sufficient to obtain precise estimates.
- Bayesian solution to VARs: shrinkage of posterior estimates towards prior choices.
- The literature on Bayesian VARs allows us to work out Big Data issues and add interesting features such as multivariate stochastic volatility and time-varying parameters.

Review the Concepts and Proofs

Bayesian VAR Models:

1. Why is shrinkage particularly important for VAR models?
2. What are the marginal distributions of a Normal-inverse Wishart distribution?
3. Explain the specification of the Minnesota prior.
4. Explain the fictitious sample interpretation of the natural conjugate prior for normally distributed VAR models.
5. In what respect in the natural conjugate prior more restrictive than the independent Normal-Wishart prior? Give an example.
6. How can you find a symmetric 90 percent H step ahead forecast interval for a VAR model that is estimated (a) with a natural conjugate prior and (b) with an independent Normal-Wishart prior?

1. Consider the M-dimensional VAR(p) model

$$y_t = a_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \qquad t = 1, \ldots, T,$$

with $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$ and $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$.

(a) Show that the VAR(p) model can be cast in *matrix representation*

$$Y = XA + E$$

and in *vector representation*

$$y = \mathbf{X}\alpha + \varepsilon.$$

Find the dimensions of the above vectors and matrices.

Solution to Exercise 1 (a)

Transpose the VAR(p) model to obtain

$$y_t' = a_0' + y_{t-1}'A_1' + \cdots + y_{t-p}'A_p' + \varepsilon_t' = [1, y_{t-1}', \ldots, y_{t-p}'] \begin{pmatrix} a_0' \\ A_1' \\ \vdots \\ A_p' \end{pmatrix} + \varepsilon_t' = x_t A + \varepsilon_t'$$

with the $(pM+1) \times M$ matrix $A = [a_0, A_1, \ldots, A_p]'$ and the $1 \times (pM+1)$ vector $x_t = [1, y_{t-1}', \ldots, y_{t-p}']$. Stacking all observations $t = 1, \ldots, T$ yields the *matrix representation*

$$Y = XA + E,$$

where $Y = [y_1, \ldots, y_T]'$ is $T \times M$, $X = [x_1', \ldots, x_T']'$ is $T \times (pM+1)$, and $E = [\varepsilon_1, \ldots, \varepsilon_T]'$ is $T \times M$.

Note that E is not a vector which is why it has a matrix-variate normal distribution.

Solution to Exercise 1 (a)

To arrive at a vector-variate normal distribution (the one we know), we vectorize the system:

$$\text{vec}(Y) = \text{vec}(XA) + \text{vec}(E) = (I_M \otimes X)\,\text{vec}(A) + \text{vec}(E),$$

where we use the rule $\text{vec}(QPR) = (R' \otimes Q)\,\text{vec}(P)$ for matrices Q, P, and R of appropriate dimensions. Defining the $TM \times 1$ vector $y = \text{vec}(Y)$, the $TM \times (pM^2 + M)$ matrix $\mathbf{X} = I_M \otimes X$, the $(pM^2 + M) \times 1$ vector $\alpha = \text{vec}(A)$, and the $TM \times 1$ vector $\varepsilon = \text{vec}(E)$ yields the *vector representation*

$$y = \mathbf{X}\alpha + \varepsilon.$$

1. **Consider the M-dimensional VAR(p) model**

$$y_t = a_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \qquad t = 1, \ldots, T,$$

with $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$ **and** $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$ **for** $t \neq s$**.**

   (b) Find the distribution of ε.

Note that ε is the $TM \times 1$ vector

$$\varepsilon = \mathrm{vec}(E) = [\varepsilon_{11}, \ldots, \varepsilon_{1T}, \varepsilon_{21}, \ldots, \varepsilon_{2T}, \ldots \ldots, \varepsilon_{M1}, \ldots, \varepsilon_{MT}]'.$$

Defining $\tilde{\varepsilon}_i = [\varepsilon_{i1}, \ldots, \varepsilon_{iT}]'$ as the vector of all T disturbances of equation $i = 1, \ldots, M$ of the VAR, we can write equivalently

$$\varepsilon = [\tilde{\varepsilon}_1', \ldots, \tilde{\varepsilon}_M']'.$$

Since each element of ε is normally distributed, ε is also normally distributed. It has mean zero because each element has mean zero.

Solution to Exercise 1 (b)

To find the variance, first note that disturbances from different time periods are uncorrelated which is why

$$\mathrm{E}(\tilde{\varepsilon}_i \tilde{\varepsilon}_j') = \sigma_{ij} I_T$$

where σ_{ij} is the row i, column j element of Σ. Now,

$$\mathrm{Var}(\varepsilon) = \mathrm{E}\left[\begin{pmatrix} \tilde{\varepsilon}_1 \\ \vdots \\ \tilde{\varepsilon}_M \end{pmatrix} [\tilde{\varepsilon}_1', \dots, \tilde{\varepsilon}_M']\right] = \mathrm{E}\begin{bmatrix} \tilde{\varepsilon}_1\tilde{\varepsilon}_1' & \tilde{\varepsilon}_1\tilde{\varepsilon}_2' & \cdots & \tilde{\varepsilon}_1\tilde{\varepsilon}_M' \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\varepsilon}_M\tilde{\varepsilon}_1' & \tilde{\varepsilon}_M\tilde{\varepsilon}_2' & \cdots & \tilde{\varepsilon}_M\tilde{\varepsilon}_M' \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} I_T & \sigma_{12} I_T & \cdots & \sigma_{1M} I_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1} I_T & \sigma_{M2} I_T & \cdots & \sigma_{MM} I_T \end{bmatrix} = \Sigma \otimes I_T.$$

Therefore, $\varepsilon \sim \mathcal{N}(0, \Sigma \otimes I_T)$.

1. Consider the M-dimensional VAR(p) model

$$y_t = a_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \qquad t = 1, \ldots, T,$$

with $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$ and $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$.

(c) Show that the joint pdf of y (conditional on p pre-sample observations $y_0, \ldots, y_{1-p}$) can be written as

$$f(y|\alpha, \Sigma) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}(y - \mathbf{X}\alpha)'(\Sigma^{-1} \otimes I_T)(y - \mathbf{X}\alpha)\right\}.$$

To this end, use an appropriate recursive factorization of the joint distribution in conditionals and marginals.

Solution to Exercise 1 (c)

To obtain the joint pdf of $y = \text{vec}([y_1, \ldots, y_T]')$, we use the conditional-marginal factorization

$$
\begin{aligned}
f(y|y_{1-p}, \ldots, y_0) &= f(y_T|\mathcal{I}_{T-1}) \cdot f(y_{T-1}|\mathcal{I}_{T-2}) \cdot \ldots \cdot f(y_1|\mathcal{I}_0) \\
&= \prod_{t=1}^{T} f(y_t|\mathcal{I}_{t-1}) \qquad (1)
\end{aligned}
$$

where $\mathcal{I}_{t-1}$ is the information set of period $t-1$.

Solution to Exercise 1 (c)

To find $f(y_t|\mathcal{I}_{t-1})$, note that the conditional mean of y_t given the past is

$$\mathrm{E}(y_t|\mathcal{I}_{t-1}) = \mathrm{E}(y_t|y_{t-1}, \dots, y_{t-p}) = a_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} = A' x_t'$$

and the conditional variance is

$$\mathrm{Var}(y_t|\mathcal{I}_{t-1}) = \mathrm{Var}(y_t|y_{t-1}, \dots, y_{t-p}) = \mathrm{Var}(\varepsilon_t) = \Sigma.$$

Also note that y_t is, conditional on the past, a linear function of the normally distributed random vector ε_t. Hence, it is also conditionally normally distributed. Taken together, we have found that $y_t|\mathcal{I}_{t-1} \sim \mathcal{N}(A' x_t', \Sigma)$ with pdf

$$f(y_t|\mathcal{I}_{t-1}) \propto |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(y_t - A' x_t')' \Sigma^{-1} (y_t - A' x_t')\right]. \tag{2}$$

Solution to Exercise 1 (c)

Substituting (2) into (1) yields

$$
\begin{aligned}
f(y|\alpha,\Sigma) &\propto \prod_{t=1}^{T} \left\{ |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(y_t - A'x_t')'\Sigma^{-1}(y_t - A'x_t') \right] \right\} \\
&\propto |\Sigma|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}\sum_{t=1}^{T}(y_t - A'x_t')'\Sigma^{-1}(y_t - A'x_t') \right]. \qquad (3)
\end{aligned}
$$

where we neglect $y_0, \ldots, y_{1-p}$ in the condition set but add $\alpha = \text{vec}(A)$ and Σ to stress the dependence on these parameters.

Solution to Exercise 1 (c)

It is convenient to express the joint pdf in a form without a summation. To this end, note that

$$\sum_{t=1}^{T}(y_t - A'x_t')'\Sigma^{-1}(y_t - A'x_t') \;=\; \sum_{t=1}^{T}\operatorname{tr}\left[(y_t - A'x_t')'\Sigma^{-1}(y_t - A'x_t')\right]$$

because it is scalar. Using the rule $\operatorname{tr}(PQR) = \operatorname{tr}(RPQ)$ for matrices P, Q, R of appropriate dimensions, we obtain

$$\begin{aligned}\sum_{t=1}^{T}(y_t - A'x_t')'\Sigma^{-1}(y_t - A'x_t') &= \sum_{t=1}^{T}\operatorname{tr}\left[(y_t - A'x_t')(y_t - A'x_t')'\Sigma^{-1}\right]\\ &= \operatorname{tr}\left[\sum_{t=1}^{T}(y_t - A'x_t')(y_t - A'x_t')'\Sigma^{-1}\right].\end{aligned}$$

Now note that $\sum_{t=1}^{T}(y_t - A'x_t')(y_t - A'x_t')' = (Y - XA)'(Y - XA)$ which allows us to write

$$\operatorname{tr}\left[\sum_{t=1}^{T}(y_t - A'x_t')(y_t - A'x_t')'\Sigma^{-1}\right] = \operatorname{tr}\left[(Y - XA)'(Y - XA)\Sigma^{-1}\right].$$

Solution to Exercise 1 (c)

Substituting this into (3) yields

$$f(y|\alpha, \Sigma) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2} \operatorname{tr}\left[(Y - XA)'(Y - XA)\Sigma^{-1}\right]\right\}. \tag{4}$$

While (4) is a perfectly sufficient solution to this question, is convenient for subsequent questions to further transform this expression. To this end, apply the rule $\operatorname{tr}(PQR) = \operatorname{vec}(P')'(R' \otimes I) \operatorname{vec}(Q)$ for matrices P, Q, R of appropriate dimensions, which yields

$$\begin{aligned} \operatorname{tr}\left[(Y - XA)'(Y - XA)\Sigma^{-1}\right] &= \operatorname{vec}(Y - XA)'(\Sigma^{-1} \otimes I_T) \operatorname{vec}(Y - XA) \\ &= (y - \mathbf{X}\alpha)'(\Sigma^{-1} \otimes I_T)(y - \mathbf{X}\alpha). \end{aligned}$$

Substituting this into the joint pdf yields

$$f(y|\alpha, \Sigma) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}(y - \mathbf{X}\alpha)'(\Sigma^{-1} \otimes I_T)(y - \mathbf{X}\alpha)\right\}. \tag{5}$$

1. Consider the M-dimensional VAR(p) model

$$y_t = a_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \qquad t = 1, \ldots, T,$$

with $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$ **and** $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$ **for** $t \neq s$**.**

(d) (*) Find the distribution of y (conditional on p pre-sample observations $y_0, \ldots, y_{1-p}$). To this end, apply the transformation technique

$$f(y) = \left| \frac{\partial \varepsilon}{\partial y'} \right| f_\varepsilon(\varepsilon)$$

and use without proof $\left| \frac{\partial \varepsilon}{\partial y'} \right| = 1$.

(e) (*) Show that the likelihood of the VAR(p) model has a Normal-Wishart structure.

Solution to Exercise 1 (e)

To preview the results, in the following we show that likelihood

$$f(y|\alpha, \Sigma) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}(y - \mathbf{X}\alpha)'(\Sigma^{-1} \otimes I_T)(y - \mathbf{X}\alpha)\right\}. \tag{6}$$

can be interpreted as the kernel of a Normal-Wishart distribution where

$$\alpha|\Sigma \sim \mathcal{N}(\hat{\alpha}, \Sigma \otimes (X'X)^{-1})$$

and

$$\Sigma \sim iW(S, T - pM - M - 2)$$

with $S = (Y - X\hat{A})'(Y - X\hat{A})$, $\hat{A} = (X'X)^{-1}X'Y$ and $\hat{\alpha} = \text{vec}(\hat{A})$. For the definition of the Wishart and inverse Wishart distributions, see Appendix.

Solution to Exercise 1 (e)

To prove this claim is a bit tedious (which is why it is omitted in the lecture). However, in principle it works as in the simple linear regression model. We proceed as follows:

1. Conjecture that α is, conditional on Σ, normally distributed with mean m_α and variance V_α, i.e., $\alpha|\Sigma \sim \mathcal{N}(m_\alpha, V_\alpha)$. Compare the pdf of this distribution with the kernel of the likelihood function to determine m_α and V_α.
2. Factorize the likelihood function into the normal pdf and a remainder term. Calculate this remainder term.
3. Conjecture that Σ is inverse Wishart distributed with parameters S and ν, i.e., $\Sigma \sim iW(S, \nu)$. Compare the pdf of this distribution and the remainder term to determine S and ν.

2. **Consider the VAR(p) model (1). Find the posterior distribution of α and Σ if you use the improper prior $f(\alpha, \Sigma) = |\Sigma|^{-\frac{M+\phi}{2}}$.**

3. **Consider the VAR(p) model (1). Find the posterior distribution of α if you use the Minnesota prior $\Sigma = \hat{\Sigma}$ and $\alpha \sim \mathcal{N}(\underline{\alpha}, \underline{V}_M)$.**

4. **Consider the VAR(p) model (1). Find the posterior distribution of α and Σ if you use the natural conjugate prior $\Sigma \sim iW(\underline{S}, \underline{\nu})$ and $\alpha|\Sigma \sim \mathcal{N}(\underline{\alpha}, \Sigma \otimes \underline{V})$.**

5. **Consider the VAR(p) model (1). Show that the following two estimation procedures yield the same posterior distributions of α and Σ: Applying the natural conjugate prior $\Sigma \sim iW(\underline{S}, \underline{\nu})$ and $\alpha|\Sigma \sim \mathcal{N}(\underline{\alpha}, \Sigma \otimes \underline{V})$ to the data Y and X, or applying the diffuse prior $f(\alpha, \Sigma) \propto |\Sigma|^{-\frac{M+1}{2}}$ to the augmented data**

$$Y_* = \begin{pmatrix} Y \\ Y_0 \end{pmatrix}, \quad X_* = \begin{pmatrix} X \\ X_0 \end{pmatrix}, \quad E_* = \begin{pmatrix} E \\ E_0 \end{pmatrix}$$

where Y_0 and X_0 contain T_0 fictitious observations chosen to satisfy the conditions (C.1) $\underline{V} = (X_0' X_0)^{-1}$, (C.2) $\underline{\alpha} = \text{vec}(\underline{A}) = \text{vec}(\hat{A}_0) \equiv \text{vec}[(X_0' X_0)^{-1} X_0' Y_0]$, and (C.3) $\underline{S} = (Y_0 - X_0 \hat{A}_0)'(Y_0 - X_0 \hat{A}_0)$.

Solution to Exercise 2: the improper prior

The joint prior pdf is

$$f(\alpha, \Sigma) \propto |\Sigma|^{-\frac{M+\phi}{2}}. \tag{7}$$

The likelihood is

$$f(y|\alpha, \Sigma) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}(y - \mathbf{X}\alpha)'(\Sigma^{-1} \otimes I_T)(y - \mathbf{X}\alpha)\right\}. \tag{8}$$

Multiplying the prior (7) with the likelihood (8) yields the posterior:

$$\begin{aligned} f(\alpha, \Sigma|y) &\propto f(y|\alpha, \Sigma) f(\alpha, \Sigma) \\ &\propto |\Sigma|^{-\frac{T+M+\phi}{2}} \exp\left\{-\frac{1}{2} \underbrace{\left[(y - \mathbf{X}\alpha)'(\Sigma^{-1} \otimes I_T)(y - \mathbf{X}\alpha)\right]}_{\equiv \kappa}\right\}. \end{aligned} \tag{9}$$

This loosely resembles the product of a normal and an inverse Wishart distribution.

Solution to Exercise 2: the improper prior

To further work this out, expand κ so that the quadratic form in α becomes clearer (also substitute $\mathbf{X} = (I_M \otimes X)$):

$$\kappa = y'(\Sigma^{-1} \otimes I_T)y - 2y'(\Sigma^{-1} \otimes I_T)\mathbf{X}\alpha + \alpha'\mathbf{X}'(\Sigma^{-1} \otimes I_T)\mathbf{X}\alpha$$

$$= y'(\Sigma^{-1} \otimes I_T)y - 2y'(\Sigma^{-1} \otimes X)\alpha + \alpha'(\Sigma^{-1} \otimes X'X)\alpha. \tag{10}$$

Substituting this back into (9) and separating into two exponential functions (one that contains α and Σ and one that contains only Σ) yields

$$f(\alpha, \Sigma|y) \propto |\Sigma|^{-\frac{T+M+\phi}{2}} \exp\left\{-\frac{1}{2}\left[y'(\Sigma^{-1} \otimes I_T)y\right]\right\}$$
$$\times \exp\left\{-\frac{1}{2}\left[\alpha'(\Sigma^{-1} \otimes X'X)\alpha - 2y'(\Sigma^{-1} \otimes X)\alpha\right]\right\}. \tag{11}$$

Solution to Exercise 2: the improper prior

Let us conjecture that (11) can be written as $f(\alpha, \Sigma|y) = f(\alpha|y, \Sigma)f(\Sigma|y)$, where $f(\alpha|y, \Sigma)$ is a normal pdf with posterior mean $\bar{\alpha}$ and posterior variance matrix $\Sigma \otimes \bar{V}$ that need to be determined. This pdf is

$$f(\alpha|y, \Sigma) \propto |\Sigma \otimes \bar{V}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\alpha - \bar{\alpha})'(\Sigma^{-1} \otimes \bar{V}^{-1})(\alpha - \bar{\alpha})\right]$$

$$\begin{aligned} &\propto |\Sigma|^{-\frac{Mp+1}{2}} |\bar{V}|^{-\frac{M}{2}} \\ &\times \exp\left\{-\frac{1}{2}\left[\alpha'(\Sigma^{-1} \otimes \bar{V}^{-1})\alpha - 2\bar{\alpha}'(\Sigma^{-1} \otimes \bar{V}^{-1})\alpha + \bar{\alpha}'(\Sigma^{-1} \otimes \bar{V}^{-1})\bar{\alpha}\right]\right\}. \end{aligned} \tag{12}$$

Solution to Exercise 2: the improper prior

Comparison of (11) and (12) shows that $\Sigma^{-1} \otimes \bar{V}^{-1} = \Sigma^{-1} \otimes X'X$ and thus

$$\bar{V} = (X'X)^{-1}. \tag{13}$$

In addition, $\bar{\alpha}'(\Sigma^{-1} \otimes \bar{V}^{-1}) = y'(\Sigma^{-1} \otimes X)$ and thus

$$\begin{aligned} \bar{\alpha}' &= y'(\Sigma^{-1} \otimes X)(\Sigma \otimes \bar{V}) = y'(I_M \otimes X\bar{V}) \\ \Rightarrow \bar{\alpha} &= (I_M \otimes X'\bar{V})y = (I_M \otimes X'(X'X)^{-1})y \end{aligned} \tag{14}$$

$$\tag{15}$$

or, equivalently,

$$\bar{A} = (X'X)^{-1}X'Y = \hat{A}. \tag{16}$$

Hence, we have found that $\alpha|\Sigma, y \sim \mathcal{N}(\text{vec}(\hat{A}), \Sigma \otimes (X'X)^{-1})$.

Solution to Exercise 2: the improper prior

In a final step, let us find

$$f(\Sigma|y) = \frac{f(\alpha, \Sigma|y)}{f(\alpha|y, \Sigma)}.$$

Substituting (11) and (12) yields

$$f(\Sigma|y) \propto \frac{|\Sigma|^{-\frac{T+M+\phi}{2}} \exp\left\{-\frac{1}{2}\left[y'(\Sigma^{-1} \otimes I_T)y\right]\right\}}{|\Sigma|^{-\frac{Mp+1}{2}} |\bar{V}|^{-\frac{M}{2}}} \times \frac{\exp\left\{-\frac{1}{2}\left[\alpha'(\Sigma^{-1} \otimes X'X)\alpha - 2y'(\Sigma^{-1} \otimes X)\alpha\right]\right\}}{\exp\left\{-\frac{1}{2}\left[\alpha'(\Sigma^{-1} \otimes \bar{V}^{-1})\alpha - 2\bar{\alpha}'(\Sigma^{-1} \otimes \bar{V}^{-1})\alpha + \bar{\alpha}'(\Sigma^{-1} \otimes \bar{V}^{-1})\bar{\alpha}\right]\right\}}$$

Note that, by construction of $\bar{\alpha}$ and $\bar{V}$, the first two additive parts (the one that is quadratic in α and the one that is linear in α) in the exponential functions are identical and thus cancel out. This yields

$$f(\Sigma|y) \propto \frac{|\Sigma|^{-\frac{T+M+\phi}{2}} \exp\left\{-\frac{1}{2}\left[y'(\Sigma^{-1} \otimes I_T)y\right]\right\}}{|\Sigma|^{-\frac{Mp+1}{2}} |\bar{V}|^{-\frac{M}{2}} \exp\left\{-\frac{1}{2}\left[\bar{\alpha}'(\Sigma^{-1} \otimes \bar{V}^{-1})\bar{\alpha}\right]\right\}}. \tag{17}$$

Solution to Exercise 2: the improper prior

Leaving out $\bar{V}$ which is not a function of α or Σ, and noting that $y'(\Sigma^{-1} \otimes I_T)y = \text{tr}(Y'Y\Sigma^{-1})$ and $\bar{\alpha}'(\Sigma^{-1} \otimes \bar{V}^{-1})\bar{\alpha} = \text{tr}(\bar{A}'\bar{V}^{-1}\bar{A}\Sigma^{-1})$, this can be simplified to

$$f(\Sigma|y) \propto |\Sigma|^{-\frac{T+M+\phi-pM-1}{2}} \exp\left\{-\frac{1}{2}\text{tr}\left[(Y'Y - \bar{A}'\bar{V}^{-1}\bar{A})\Sigma^{-1}\right]\right\}. \tag{18}$$

This is the kernel of an inverse Wishart pdf with parameters

$$\bar{\nu} = T + \phi - pM - 2. \tag{19}$$

and (using $M_X = I - X(X'X)^{-1}X'$)

$$\begin{aligned}\bar{S} &= Y'Y - \bar{A}'\bar{V}^{-1}\bar{A} = Y'Y - Y'X(X'X)^{-1}X'X(X'X)^{-1}X'Y \\ &= Y'M_XY = (Y - X\hat{A})'(Y - X\hat{A}).\end{aligned}$$

Hence, we have found that $\Sigma|y \sim iW((Y - X\hat{A})'(Y - X\hat{A}), T + \phi - pM - 2)$.

Solution to Exercise 3: the Minnesota prior

The Minnesota prior sets $\Sigma = \hat{\Sigma}$. Hence, we only need to estimate α.
The prior $\alpha \sim \mathcal{N}(\underline{\alpha}, \underline{V}_M)$ implies the pdf kernel

$$\begin{aligned} f(\alpha) &\propto \exp\left[-\frac{1}{2}(\alpha - \underline{\alpha})' \underline{V}_M^{-1} (\alpha - \underline{\alpha})\right] \\ &\propto \exp\left\{-\frac{1}{2}\left[\alpha' \underline{V}_M^{-1} \alpha - 2\underline{\alpha}' \underline{V}_M^{-1} \alpha + \underline{\alpha}' \underline{V}_M^{-1} \underline{\alpha}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\alpha' \underline{V}_M^{-1} \alpha - 2\underline{\alpha}' \underline{V}_M^{-1} \alpha\right]\right\}, \end{aligned} \tag{20}$$

where we leave out any factor that is not a function of α.

Solution to Exercise 3: the Minnesota prior

Since $\Sigma = \hat{\Sigma}$ is treated as known, the likelihood reduces to

$$\begin{aligned} f(y|\alpha) &\propto |\hat{\Sigma}|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}(y - \mathbf{X}\alpha)'(\hat{\Sigma}^{-1} \otimes I_T)(y - \mathbf{X}\alpha)\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[y'(\hat{\Sigma}^{-1} \otimes I_T)y - 2y'(\hat{\Sigma}^{-1} \otimes I_T)\mathbf{X}\alpha + \alpha'\mathbf{X}'(\hat{\Sigma}^{-1} \otimes I_T)\mathbf{X}\alpha\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\alpha'\mathbf{X}'(\hat{\Sigma}^{-1} \otimes I_T)\mathbf{X}\alpha - 2y'(\hat{\Sigma}^{-1} \otimes I_T)\mathbf{X}\alpha\right]\right\} \end{aligned}$$

where we leave out again any factor that is not a function of α. Substituting $\mathbf{X} = (I_M \otimes X)$ yields

$$\begin{aligned} f(y|\alpha) &\propto \exp\left\{-\frac{1}{2}\left[\alpha'(I_M \otimes X')(\hat{\Sigma}^{-1} \otimes I_T)(I_M \otimes X)\alpha - 2y'(\hat{\Sigma}^{-1} \otimes I_T)(I_M \otimes X)\alpha\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\alpha'(\hat{\Sigma}^{-1} \otimes X'X)\alpha - 2y'(\hat{\Sigma}^{-1} \otimes X)\alpha\right]\right\} \end{aligned} \tag{21}$$

Solution to Exercise 3: the Minnesota prior

Combining prior (20) and likelihood (21) yields the posterior

$$\begin{aligned} f(\alpha|y) &\propto f(y|\alpha)f(\alpha) \\ &\propto \exp\left\{-\frac{1}{2}\left[\alpha'(\hat{\Sigma}^{-1}\otimes X'X)\alpha + \alpha'\underline{V}_M^{-1}\alpha - 2y'(\hat{\Sigma}^{-1}\otimes X)\alpha - 2\underline{\alpha}'\underline{V}_M^{-1}\alpha\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\alpha'(\hat{\Sigma}^{-1}\otimes X'X + \underline{V}_M^{-1})\alpha - 2\left(y'(\hat{\Sigma}^{-1}\otimes X) + \underline{\alpha}'\underline{V}_M^{-1}\right)\alpha\right]\right\}. \end{aligned} \tag{22}$$

Since this resembles the kernel of a normal distribution, let us conjecture that the posterior is a normal with mean $\bar{\alpha}$ and variance matrix $\bar{V}_M$ that have to be determined. This distribution has pdf

$$\begin{aligned} f(\alpha|y) &\propto \exp\left\{-\frac{1}{2}(\alpha-\bar{\alpha})'\bar{V}_M^{-1}(\alpha-\bar{\alpha})\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\alpha'\bar{V}_M^{-1}\alpha - 2\bar{\alpha}'\bar{V}_M^{-1}\alpha + \bar{\alpha}'\bar{V}_M^{-1}\bar{\alpha}\right]\right\} \\ &\propto \exp\left\{-\frac{1}{2}\left[\alpha'\bar{V}_M^{-1}\alpha - 2\bar{\alpha}'\bar{V}_M^{-1}\alpha\right]\right\}. \end{aligned} \tag{23}$$

Solution to Exercise 3: the Minnesota prior

Comparison of (22) and (23) shows that $\bar{V}_M^{-1} = \hat{\Sigma}^{-1} \otimes X'X + \underline{V}_M^{-1}$ and thus

$$\bar{V}_M = (\hat{\Sigma}^{-1} \otimes X'X + \underline{V}_M^{-1})^{-1}. \tag{24}$$

The comparison also shows that $\bar{\alpha}' \bar{V}_M^{-1} = y'(\hat{\Sigma}^{-1} \otimes X) + \underline{\alpha}' \underline{V}_M^{-1}$ and thus

$$\bar{\alpha} = \bar{V}_M[(\hat{\Sigma}^{-1} \otimes X')y + \underline{V}_M^{-1} \underline{\alpha}].$$

Using $(\hat{\Sigma}^{-1} \otimes X')y = (\hat{\Sigma}^{-1} \otimes X'X)(I_M \otimes (X'X)^{-1}X')y = (\hat{\Sigma}^{-1} \otimes X'X)\,\mathrm{vec}[(X'X)^{-1}X'Y] = (\hat{\Sigma}^{-1} \otimes X'X)\hat{\alpha}$,

it can also be written as

$$\bar{\alpha} = \bar{V}_M[(\hat{\Sigma}^{-1} \otimes X'X)\hat{\alpha} + \underline{V}_M^{-1} \underline{\alpha}], \tag{25}$$

which demonstrates that $\bar{\alpha}$ is a weighted average of the OLS estimator and the prior mean.

Taken together, we have found that the posterior distribution of α is normal with mean $\bar{\alpha}$ given in (25) and variance $\bar{V}_M$ given in (24).

Computer-Based Exercise: BVAR with the Minnesota prior

Use the data set **US_macrodata.xlsx** to estimate a Bayesian VAR with the Minnesota prior, where the general M-dimensional VAR(p) model is considered:

$$y_t = a_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \qquad t = 1, \ldots, T,$$

with $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$ and $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$.

Appendix 1: The Wishart distribution

A good reference is Steven W. Nydick (2012), The Wishart and Inverse Wishart Distributions, downloadable here.

Let H be an $M \times M$ random matrix that follows a Wishart distribution with parameters S and ν, $H \sim W(S, \nu)$. Then it has pdf

$$f_W(H|S,\nu) = c_W^{-1} |S|^{-\frac{\nu}{2}} |H|^{\frac{\nu - M - 1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}(S^{-1}H)\right],$$

where $c_W = 2^{\frac{\nu M}{2}} \pi^{\frac{M(M-1)}{4}} \prod_{i=1}^{M} \Gamma(\frac{\nu+1-i}{2})$ is an integration constant, $\nu > M - 1$ is a scalar parameter, and S is an $M \times M$ symmetric and positive definite scale matrix. The expectation is

$$\mathrm{E}(H) = \nu S.$$

The Wishart distribution is often used as a prior for the precision matrix (=inverse of the variance matrix).

Appendix 1: The inverse Wishart distribution

Let Σ be an $M \times M$ random matrix that follows an inverse Wishart distribution with parameters Ψ and δ, $\Sigma \sim iW(\Psi, \delta)$. Then it has pdf

$$f_{iW}(\Sigma|\Psi, \delta) = c_{iW}^{-1} |\Psi|^{\frac{\delta}{2}} |\Sigma|^{-\frac{\delta+M+1}{2}} \exp\left[-\frac{1}{2}\,\mathrm{tr}(\Psi\Sigma^{-1})\right],$$

where $c_{iW} = 2^{\frac{\delta M}{2}} \pi^{\frac{M(M-1)}{4}} \prod_{i=1}^{M} \Gamma(\frac{\delta+1-i}{2})$ is an integration constant, $\delta > M - 1$ is a scalar parameter, and Ψ is an $M \times M$ symmetric and positive definite scale matrix. The expectation is

$$\mathrm{E}(\Sigma) = \Psi/(\delta - M - 1).$$

The inverse Wishart distribution is often used as a prior for the variance matrix.

Relationship between Wishart and inverse Wishart distribution:

let $H \sim W(S, \nu)$, then $\Sigma \equiv H^{-1} \sim iW(S^{-1}, \nu)$.

Appendix 2: The matrix variate t distribution

The $p \times q$ random matrix X has matrix variate t distribution, $X \sim MT(\mu, V, S, \nu)$, if it has pdf

$$f(X) = c_{MT}^{-1} |S|^{\frac{\nu}{2}} |V|^{-\frac{q}{2}} |S + (X - \mu)' V^{-1} (X - \mu)|^{-\frac{\nu+p}{2}}$$

where μ is a $p \times q$ symmetric and positive definite matrix, S is a symmetric and positive definite $q \times q$ matrix, V is a $p \times p$ matrix, $\nu > q - 1$ is a scalar and

$$c_{MT} = \pi^{\frac{pq}{2}} \prod_{i=1}^{q} \frac{\Gamma(\frac{\nu+1-i}{2})}{\Gamma(\frac{\nu+p+1-i}{2})}.$$

The matrix variate t distribution has mean

$$\mathrm{E}(X) = \mu, \qquad \nu > 1,$$

and variance

$$\mathrm{Var}[\mathrm{vec}(X)] = \frac{1}{\nu - q - 1} S \otimes V, \qquad \nu > q + 1.$$

Appendix 2: The matrix variate t distribution

A single element X_{ij} has mean

$$\mathrm{E}(X_{ij}) = \mu_{ij}, \qquad \nu > 1,$$

and variance

$$\mathrm{Var}(X_{ij}) = \frac{1}{\nu - q - 1} V_{ii} S_{jj}, \qquad \nu > q + 1.$$

Its marginal distribution is the non-standardized t distribution, $X_{ij} \sim t(\mu_{ij}, \sigma_{ij}^2, \nu - q + 1)$, where $\sigma_{ij}^2 = V_{ii} S_{jj} / (\nu - q + 1)$. The standardized element

$$T_{ij} = \frac{X_{ij} - \mu_{ij}}{\sigma_{ij}} = \frac{X_{ij} - \mu_{ij}}{\sqrt{V_{ii} S_{jj} / (\nu - q + 1)}}$$

has student t distribution, i.e., $T_{ij} \sim t(\nu - q + 1)$. In a Bayesian context, the matrix variate t distribution appears as a marginal distribution of a normal-inverse Wishart distribution. Suppose

$$\mathrm{vec}(X) | \Sigma \sim \mathcal{N}(\mathrm{vec}(\mu), \Sigma \otimes V), \quad \Sigma \sim iW(S, \nu).$$

Then

$$X \sim MT(\mu, V, S, \nu).$$

Appendix 3: Rules for matrices

Suppose all matrices P, Q, R, and S are of appropriate dimensions. Then

1. $\text{vec}(QPR) = (R' \otimes Q)\,\text{vec}(P)$
2. $\text{tr}(PQR) = \text{tr}(RPQ)$
3. $\text{tr}(PQR) = \text{vec}(P')'(C' \otimes I)\,\text{vec}(Q)$
4. $\text{tr}(PQRS) = \text{vec}(P')'(S' \otimes Q)\,\text{vec}(R)$
5. $|P \otimes Q| = |P|^n|Q|^m$ for an $m \times m$ matrix P and an $n \times n$ matrix Q