Lecture 6: Bayesian VAR Models

Derivation of the Posterior Distributions

1 The VAR model

The VAR(p) model is

$$y_t = a_0 + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0, \Sigma).$$

For ease of notation, the transposed VAR(p) model

$$y_t' = a_0' + y_{t-1}' A_1' + \cdots + y_{t-p}' A_p' + \varepsilon_t'$$

can be written as

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

with the $(pM+1) \times M$ matrix A and the $1 \times (pM+1)$ vector x_t so defined. 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. To arrive at a vector normal distribution (the one we know), we vectorize the system:

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

where we use the rule $\operatorname{vec}(QPR) = (R' \otimes Q) \operatorname{vec}(P)$ for matrices Q, P, and R of appropriate dimensions. Defining $y = \operatorname{vec}(Y)$, $\mathbf{X} = I_M \otimes X$, $\alpha = \operatorname{vec}(A)$, and $\varepsilon = \operatorname{vec}(E)$ yields the *vector representation*

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

where ε is the $TM \times 1$ vector

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

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

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

It has mean zero and variance matrix

$$\begin{aligned}\mathrm{Var}(\varepsilon) &= \mathrm{E}\left[\begin{pmatrix}\tilde{\varepsilon}_1 \\ \vdots \\ \tilde{\varepsilon}_M\end{pmatrix}[\tilde{\varepsilon}_1', \ldots, \tilde{\varepsilon}_M']\right] = \mathrm{E}\begin{bmatrix}\tilde{\varepsilon}_1\tilde{\varepsilon}_1' & \tilde{\varepsilon}_1\tilde{\varepsilon}_2' & \cdots & \tilde{\varepsilon}_1\tilde{\varepsilon}_M' \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\varepsilon}_M\tilde{\varepsilon}_1' & \tilde{\varepsilon}_M\tilde{\varepsilon}_2' & \cdots & \tilde{\varepsilon}_M\tilde{\varepsilon}_M'\end{bmatrix} \\ &= \begin{bmatrix}\sigma_{11}I_T & \sigma_{12}I_T & \cdots & \sigma_{1M}I_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1}I_T & \sigma_{M2}I_T & \cdots & \sigma_{MM}I_T\end{bmatrix} = \Sigma \otimes I_T,\end{aligned}$$

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

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

2 Likelihood

In the following, we show that the likelihood of the VAR model is, conditional on the parameters α and Σ and on p pre-sample values $y_0, \ldots, y_{1-p}$, proportional to

$$\begin{aligned}&f(y_1, \ldots, y_T|\alpha, \Sigma, y_0, \ldots, y_{1-p}) \\ &\quad = |\Sigma|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}(y - (I_M \otimes X)\alpha)'(\Sigma^{-1} \otimes I_T)(y - (I_M \otimes X)\alpha)\right]. \qquad (1)\end{aligned}$$

For notational convenience, we will henceforth neglect $y_0, \ldots, y_{1-p}$ in the condition set of the likelihood.

There are at least two ways to show that (1) is the likelihood. For completeness, we present both.

2.1 Derivation from recursive conditional-marginal factorization

Note that the conditional mean of y_t given p past observations is

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

and the conditional variance is

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

Since y_t is, conditional on the past, a linear function of the normally distributed random vector ε_t, it is conditionally normally distributed, $y_t|y_{t-1},\dots,y_{t-p} \sim \mathcal{N}(A'x_t',\Sigma)$ with pdf

$$f(y_t|y_{t-1},\dots,y_{t-p}) \propto |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(y_t - A'x_t')'\Sigma^{-1}(y_t - A'x_t')\right]. \tag{2}$$

To obtain the joint pdf of $y_1,\dots,y_T$, we use the conditional-marginal factorization

$$\begin{aligned} f(y_1,\dots,y_T|y_{1-p},\dots,y_0) &= f(y_T|y_{T-1},\dots,y_{T-p}) \cdot f(y_{T-1}|y_{T-2},\dots,y_{T-p-1}) \cdot \\ &\quad \dots \cdot f(y_1|y_0,\dots,y_{1-p}) \\ &= \prod_{t=1}^{T} f(y_t|y_{t-1},\dots,y_{t-p}). \end{aligned}$$

Substituting (2) and neglecting $y_0,\dots,y_{1-p}$ in the condition set yields

$$\begin{aligned} f(y_1,\dots,y_T) &\propto \prod_{t=1}^{T} \left\{ |\Sigma|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(y_t - A'x_t')'\Sigma^{-1}(y_t - A'x_t')\right]\right\} \\ &\propto |\Sigma|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}\sum_{t=1}^{T}(y_t - A'x_t')'\Sigma^{-1}(y_t - A'x_t')\right]. \end{aligned} \tag{3}$$

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

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

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

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

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

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

Substituting this into (3) yields

$$f(y_1, \ldots, y_T) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[(Y - XA)'(Y - XA)\Sigma^{-1}\right]\right\},$$

which is essentially the expression shown in Kadiyala and Karlsson (1997, p. 101).

To obtain (1), we 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

$$f(y_1, \ldots, y_T) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}(y - \mathbf{X}\alpha)'(\Sigma^{-1} \otimes I_T)(y - \mathbf{X}\alpha)\right\},$$

which is the same as (1) because $\mathbf{X} = (I_M \otimes X)$.

2.2 Derivation from distribution of ε

Alternatively, we can start from the finding that

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

with pdf

$$f_\varepsilon(\varepsilon) \propto |\Sigma \otimes I_T|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\varepsilon'(\Sigma^{-1} \otimes I_T)\varepsilon\right\}.$$

Using the rule $|P \otimes Q| = |P|^n |Q|^m$ for an $m \times m$ matrix P and an $n \times n$ matrix Q, we obtain $|\Sigma \otimes I_T| = |\Sigma|^T |I_T|^M = |\Sigma|^T$ and thus

$$f_\varepsilon(\varepsilon) \propto |\Sigma|^{-\frac{T}{2}} \exp\left\{-\frac{1}{2}\varepsilon'(\Sigma^{-1} \otimes I_T)\varepsilon\right\}.$$

To find the pdf of y, we can use the transformation technique:

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

where $\varepsilon = y - \mathbf{X}\alpha$. Using $|\partial\varepsilon/\partial y'| = 1$, we obtain

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

For the interested students: The somewhat tricky part is to find $\left|\frac{\partial \varepsilon}{\partial y'}\right|$. But it is relatively straightforward to find a related derivative if we rearrange $\varepsilon = \text{vec}(E)$ and $y = \text{vec}(Y)$ to

$$\bar{\varepsilon} = \text{vec}(E') = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix} \qquad \text{and} \qquad \bar{y} = \text{vec}(Y') = \begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix}.$$

Let us start with the first derivative of y_t with respect to ε_s. Recall that y_t is, according to the VMA representation, a linear combination of current and past disturbances:

$$y_t = \mu + I_M\varepsilon_t + C_1\varepsilon_{t-1} + C_2\varepsilon_{t-2} + \dots.$$

Hence,

$$\frac{\partial y_t}{\partial \varepsilon_s'} = \begin{cases} 0 & \text{if } s > t \\ I_M & \text{if } s = t \\ C_{t-s} & \text{if } s < t \end{cases}.$$

Now,

$$\frac{\partial \bar{y}}{\partial \bar{\varepsilon}'} = \begin{pmatrix} \frac{\partial y_1}{\partial \varepsilon_1'} & \cdots & \frac{\partial y_1}{\partial \varepsilon_T'} \\ \vdots & & \vdots \\ \frac{\partial y_T}{\partial \varepsilon_1'} & \cdots & \frac{\partial y_T}{\partial \varepsilon_T'} \end{pmatrix} = \begin{pmatrix} I_M & 0 & \cdots & 0 \\ C_1 & I_M & & \vdots \\ \vdots & & \ddots & \\ C_{T-1} & \cdots & & I_M \end{pmatrix}.$$

This shows that $\frac{\partial \bar{y}}{\partial \bar{\varepsilon}'}$ is a lower triangular matrix with ones on the main diagonal. Such a matrix has determinant one. Thus,

$$\left|\frac{\partial \bar{\varepsilon}}{\partial \bar{y}'}\right| = \left|\frac{\partial \bar{y}}{\partial \bar{\varepsilon}'}\right|^{-1} = 1^{-1} = 1.$$

Finally, we need to consider the relationship between $\left|\frac{\partial \bar{\varepsilon}}{\partial \bar{y}'}\right|$ and $\left|\frac{\partial \varepsilon}{\partial y'}\right|$. To this end, note that y and $\bar{y}$ consist of the same—but differently ordered—elements. It is possible to define a $(MT \times MT)$ dimensional commutation matrix K_{MT} which consists only of zeros and ones such that $\varepsilon = K_{MT}\bar{\varepsilon}$. Similarly, $\bar{y} = K_{TM}y$. (For a definition and properties of commutation matrices, see Lütkepohl, 2005, New Introduction to Multiple Time Series Analysis, Springer, p. 662ff.) Using the chain rule for vector differentiation,

$$\frac{\partial \varepsilon}{\partial y'} = \frac{\partial \varepsilon}{\partial \bar{\varepsilon}'}\frac{\partial \bar{\varepsilon}}{\partial \bar{y}'}\frac{\partial \bar{y}}{\partial y} = K_{MT}\frac{\partial \bar{\varepsilon}}{\partial \bar{y}'}K_{TM}.$$

Hence,

$$\left|\frac{\partial \varepsilon}{\partial y'}\right| = \left|K_{MT}\frac{\partial \bar{\varepsilon}}{\partial \bar{y}'}K_{TM}\right| = |K_{MT}|\left|\frac{\partial \bar{\varepsilon}}{\partial \bar{y}'}\right||K_{TM}| = |K_{MT}|\,|K_{TM}|\,.$$

According to Lütkepohl (2005, p. 664) $|K_{MT}| = |K_{TM}| = (-1)^{MT(M-1)(T-1)/4}$, which yields

$$\left|\frac{\partial \varepsilon}{\partial y'}\right| = (-1)^{MT(M-1)(T-1)/2} = 1,$$

where the last equality comes from the observation that for relevant cases ($M > 1$ and $T > 1$), the exponent is positive and even.

2.3 Relationship with the Normal-Wishart distribution

The likelihood (1) can be shown to be of the Normal-Wishart type which is a multivariate generalization of the Normal-Gamma distribution. Specifically, the likelihood function 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

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

where $iW(\cdot,\cdot)$ denotes an inverse Wishart distribution (see Appendix), $S = (Y - X\hat{A})'(Y - X\hat{A})$, $\hat{A} = (X'X)^{-1}X'Y$ and $\hat{\alpha} = \text{vec}(\hat{A})$. Equivalently,

$$\Sigma^{-1} \sim W(S^{-1}, T - pM - M - 2),$$

where $W(\cdot,\cdot)$ denotes a Wishart distribution (see Appendix).

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.

These three steps are shown in the following.
Step 1. Let us start from the likelihood

$$f(y_1, \ldots, y_T|\alpha, \Sigma) = |\Sigma|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}(y - (I_M \otimes X)\alpha)'(\Sigma^{-1} \otimes I_T)(y - (I_M \otimes X)\alpha)\right].$$

Interpret it as a pdf for α and Σ conditional on the data (for ease of notation, we just put y in the conditioning set) and consider a conditional-marginal factorization:

$$f(y_1, \ldots, y_T|\alpha, \Sigma) \propto f(\alpha, \Sigma|y) = f(\alpha|\Sigma, y)f(\Sigma|y).$$

If the $pM^2 + M \times 1$ vector α is conditionally normal, $\alpha|\Sigma \sim \mathcal{N}(m_\alpha, V_\alpha)$, its pdf has the kernel

$$\begin{aligned} f(\alpha|\Sigma, y) &\propto |V_\alpha|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\alpha - m_\alpha)'V_\alpha^{-1}(\alpha - m_\alpha)\right] \\ &= |V_\alpha|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}\left(\alpha'V_\alpha^{-1}\alpha - 2m_\alpha'V_\alpha^{-1}\alpha + m_\alpha'V_\alpha^{-1}m_\alpha\right)\right] \qquad (4) \end{aligned}$$

Now manipulate the argument of the exponential function of the likelihood to make it "similar" to (4):

$$\begin{aligned} &-\frac{1}{2}(y - (I_M \otimes X)\alpha)'(\Sigma^{-1} \otimes I_T)(y - (I_M \otimes X)\alpha) \\ &= -\frac{1}{2}(y' - \alpha'(I_M \otimes X'))(\Sigma^{-1} \otimes I_T)(y - (I_M \otimes X)\alpha) \\ &= -\frac{1}{2}\left[\alpha'(I_M \otimes X')(\Sigma^{-1} \otimes I_T)(I_M \otimes X)\alpha - 2y'(\Sigma^{-1} \otimes I_T)(I_M \otimes X)\alpha + y'(\Sigma^{-1} \otimes I_T)y\right] \\ &= -\frac{1}{2}\left[\alpha'(\Sigma^{-1} \otimes X'X)\alpha - 2y'(\Sigma^{-1} \otimes X)\alpha + y'(\Sigma^{-1} \otimes I_T)y\right] \qquad (5) \end{aligned}$$

Comparison of (5) with the normal kernel (4) suggests that

$$\alpha'V_\alpha^{-1}\alpha \stackrel{!}{=} \alpha'(\Sigma^{-1} \otimes X'X)\alpha$$

which implies $V_\alpha^{-1} = \Sigma^{-1} \otimes X'X$ and thus

$$V_\alpha = \Sigma \otimes (X'X)^{-1}. \tag{6}$$

The comparison also suggests that

$$2m_\alpha' V_\alpha^{-1}\alpha \stackrel{!}{=} 2y'(\Sigma^{-1} \otimes X)\alpha = 2\,\mathrm{vec}(Y)'(I_M \otimes X(X'X)^{-1})(\Sigma^{-1} \otimes X'X)\alpha$$

which implies $m_\alpha' = \mathrm{vec}(Y)'(I_M \otimes X(X'X)^{-1})$. Transposing yields

$$m_\alpha = (I_M \otimes (X'X)^{-1}X')\,\mathrm{vec}(Y) = \mathrm{vec}((X'X)^{-1}X'Y) = \mathrm{vec}(\hat{A}) = \hat{\alpha}, \tag{7}$$

where $\hat{A} = (X'X)^{-1}X'Y$ is the OLS estimator of the VAR parameter matrix A and $\hat{\alpha} = \mathrm{vec}(\hat{A})$.

Now we have determined $m_\alpha = \hat{\alpha}$ and $V_\alpha = \Sigma \otimes (X'X)^{-1}$. Thus the conditional distribution of α is

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

with pdf

$$f(\alpha|\Sigma, y) \propto |\Sigma \otimes (X'X)^{-1}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\alpha - \hat{\alpha})'(\Sigma^{-1} \otimes X'X)(\alpha - \hat{\alpha})\right]. \tag{8}$$

Step 2. In this step, we factorize the likelihood function into $f(\alpha|\Sigma, y)$ and a remainder term $f(\Sigma|y)$ which will be the kernel of the distribution for Σ:

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

As we know by now $f(\alpha|\Sigma, y)$, we find $f(\Sigma|y)$ as follows:

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

As a first simplification, recall that for $m \times m$ matrix P and a $n \times n$ matrix Q it holds that $|P \otimes Q| = |P|^n|Q|^m$. Hence, $|\Sigma \otimes (X'X)^{-1}|^{-\frac{1}{2}} = |\Sigma|^{-\frac{pM+1}{2}}|(X'X)^{-1}|^{-\frac{M}{2}}$ because Σ has dimension $M \times M$ and $(X'X)^{-1}$ has dimension $(pM+1) \times (pM+1)$. Since $|(X'X)^{-1}|^{-\frac{M}{2}}$ does not include α or Σ, it is irrelevant for the kernel and we can neglect it which yields

$$f(\Sigma|y) \propto \frac{|\Sigma|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}(y - (I_M \otimes X)\alpha)'(\Sigma^{-1} \otimes I_T)(y - (I_M \otimes X)\alpha)\right]}{|\Sigma|^{-\frac{pM+1}{2}} \exp\left[-\frac{1}{2}(\alpha - \hat{\alpha})'(\Sigma^{-1} \otimes X'X)(\alpha - \hat{\alpha})\right]}.$$

Next recall that the exponential function in the nominator can be written as

$$\exp\left\{-\frac{1}{2}\left[\alpha'(\Sigma^{-1}\otimes X'X)\alpha - 2y'(\Sigma^{-1}\otimes X)\alpha + y'(\Sigma^{-1}\otimes I_T)y\right]\right\}$$

whereas the exponential function in the denominator is

$$\begin{aligned}
&\exp\left\{-\frac{1}{2}\left[\alpha'V_\alpha^{-1}\alpha - 2m_\alpha'V_\alpha^{-1}\alpha + m_\alpha'V_\alpha^{-1}m_\alpha\right]\right\}\\
&= \exp\left\{-\frac{1}{2}\left[\alpha'(\Sigma^{-1}\otimes X'X)\alpha - 2y'(\Sigma^{-1}\otimes X)\alpha + \hat{\alpha}'(\Sigma^{-1}\otimes X'X)\hat{\alpha}\right]\right\}.
\end{aligned}$$

Thus, their ratio simplifies to

$$\exp\left\{-\frac{1}{2}\left[y'(\Sigma^{-1}\otimes I_T)y - \hat{\alpha}'(\Sigma^{-1}\otimes X'X)\hat{\alpha}\right]\right\}.$$

This yields the remainder term

$$f(\Sigma|y) \propto |\Sigma|^{-\frac{T-pM-1}{2}} \exp\left\{-\frac{1}{2}\left[y'(\Sigma^{-1}\otimes I_T)y - \hat{\alpha}'(\Sigma^{-1}\otimes X'X)\hat{\alpha}\right]\right\}. \tag{9}$$

Step 3. Finally, let us conjecture that the remainder term is the kernel of the pdf of an inverse Wishart distribution with parameters S and ν, and determine these parameters such that

$$f(\Sigma|y) \propto |\Sigma|^{-\frac{\nu+M+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}(S\Sigma^{-1})\right].$$

Comparison with (9) suggests that $T-pM-1 = \nu+M+1$ and thus $\nu = T-pM-M-2$. However, we still have to transform the argument of the exponential function. Substituting $\hat{\alpha} = (I_M\otimes(X'X)^{-1}X')y$ into (9) yields

$$\begin{aligned}
&-\frac{1}{2}\left[y'(\Sigma^{-1}\otimes I_T)y - y'(I_M\otimes X(X'X)^{-1})(\Sigma^{-1}\otimes X'X)(I_M\otimes(X'X)^{-1}X')y\right]\\
&= -\frac{1}{2}\left[y'(\Sigma^{-1}\otimes I_T)y - y'(\Sigma^{-1}\otimes X(X'X)^{-1}X')y\right]\\
&= -\frac{1}{2}\left[y'(\Sigma^{-1}\otimes I_T - \Sigma^{-1}\otimes X(X'X)^{-1}X')y\right]\\
&= -\frac{1}{2}\left[y'(\Sigma^{-1}\otimes[I_T - X(X'X)^{-1}X'])y\right]\\
&= -\frac{1}{2}\left[\operatorname{vec}(Y)'(\Sigma^{-1}\otimes M_X)\operatorname{vec}(Y)\right],
\end{aligned}$$

where we used the definition $M_X = I_T - X(X'X)^{-1}X'$. Note that for matrices of suitable dimension the rule $\mathrm{tr}(PQRS) = \mathrm{vec}(P')'(S' \otimes Q)\,\mathrm{vec}(R)$ holds. Applying it yields

$$-\frac{1}{2}\left[\mathrm{vec}(Y)'(\Sigma^{-1} \otimes M_X)\,\mathrm{vec}(Y)\right] = -\frac{1}{2}\,\mathrm{tr}(Y'M_XY\Sigma^{-1}).$$

Hence, the remainder kernel (9) can be written as

$$f(\Sigma|y) \propto |\Sigma|^{-\frac{T-pM-1}{2}} \exp\left[-\frac{1}{2}\,\mathrm{tr}(Y'M_XY\Sigma^{-1})\right]. \tag{10}$$

Comparison with the inverse Wishart kernel shows that $S = Y'M_XY$. Hence, we have shown that the remainder term is the kernel of an inverse Wishart distribution with parameters $\nu = T - pM - M - 2$ and $\Psi = Y'M_XY$. Note that S is the multivariate analogue of the residual sum of squares:

$$S = Y'M_XY = (M_XY)'(M_XY) = \hat{E}'\hat{E} = (Y - X\hat{A})'(Y - X\hat{A}).$$

3 Priors and posteriors

In the following we derive the posteriors based on three different priors; the Minnesota prior, the Normal-Wishart prior, and the independent Normal-Wishart prior. For a discussion of these priors, see, e.g., Koop and Korobilis (2010), Bayesian Multivariate Time Series Methods for Empirical Macroeconomics, Foundations and Trends in Econometrics, Vol. 3, No. 4, p. 267-358.

3.1 Minnesota prior

The Minnesota prior sets $\Sigma = \hat{\Sigma}$. Hence, we only need to estimate α. It is assumed that the prior distribution is $\alpha \sim \mathcal{N}(\underline{\alpha}, \underline{V})$ with pdf

$$f(\alpha) \propto \exp\left[-\frac{1}{2}(\alpha - \underline{\alpha})'\underline{V}^{-1}(\alpha - \underline{\alpha})\right].$$

Hence, the posterior is

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

Substituting $\mathbf{X} = (I_M \otimes X)$ yields

$$\begin{aligned}
f(\alpha|y) &\propto \exp\left\{-\frac{1}{2}\left[\alpha'(\hat{\Sigma}^{-1}\otimes X'X)\alpha + \alpha'\underline{V}^{-1}\alpha - 2y'(\hat{\Sigma}^{-1}\otimes X)\alpha - 2\underline{\alpha}'\underline{V}^{-1}\alpha\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\alpha'(\hat{\Sigma}^{-1}\otimes X'X + \underline{V}^{-1})\alpha - 2\left(y'(\hat{\Sigma}^{-1}\otimes X) + \underline{\alpha}'\underline{V}^{-1}\right)\alpha\right]\right\}. \qquad (11)
\end{aligned}$$

Let us conjecture that the posterior is a Normal distribution with mean $\bar{\alpha}$ and variance matrix $\bar{V}$. Then the pdf is

$$\begin{aligned}
f(\alpha|y) &\propto \exp\left\{-\frac{1}{2}(\alpha-\bar{\alpha})'\bar{V}^{-1}(\alpha-\bar{\alpha})\right\} = \exp\left\{-\frac{1}{2}\left[\alpha'\bar{V}^{-1}\alpha - 2\bar{\alpha}'\bar{V}^{-1}\alpha + \bar{\alpha}'\bar{V}^{-1}\bar{\alpha}\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\alpha'\bar{V}^{-1}\alpha - 2\bar{\alpha}'\bar{V}^{-1}\alpha\right]\right\}. \qquad (12)
\end{aligned}$$

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

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

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

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

3.2 Natural conjugate Normal-Wishart prior

Recall that the likelihood has a Normal-Wishart form. Hence, a Normal-Wishart prior is a natural conjugate prior. The variance matrix follows an inverse Wishart distribution, $\Sigma \sim iW(\underline{S}, \underline{\nu})$, with pdf

$$f(\Sigma) \propto |\Sigma|^{-\frac{\underline{\nu}+M+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}(\underline{S}\Sigma^{-1})\right],$$

and the mean parameters follow a conditional normal distribution, $\alpha|\Sigma \sim \mathcal{N}(\underline{\alpha}, \Sigma \otimes \underline{V})$ with pdf

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

Hence the joint prior pdf is

$$f(\alpha,\Sigma) \propto |\Sigma|^{-\frac{\underline{\nu}+Mp+M+2}{2}} \exp\left[-\frac{1}{2}(\alpha-\underline{\alpha})'(\Sigma^{-1}\otimes \underline{V}^{-1})(\alpha-\underline{\alpha})\right] \exp\left[-\frac{1}{2}\operatorname{tr}(\underline{S}\Sigma^{-1})\right].$$

Recall that the likelihood is

$$\begin{aligned} f(y|\alpha,\Sigma) \propto |\Sigma|^{-\frac{T-pM-1}{2}} &\exp\left[-\frac{1}{2}\operatorname{tr}(S\Sigma^{-1})\right] \\ &\times |\Sigma\otimes(X'X)^{-1}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\alpha-\hat{\alpha})'(\Sigma^{-1}\otimes X'X)(\alpha-\hat{\alpha})\right] \\ &\propto |\Sigma|^{-\frac{T}{2}} \exp\left[-\frac{1}{2}(\alpha-\hat{\alpha})'(\Sigma^{-1}\otimes X'X)(\alpha-\hat{\alpha})\right] \exp\left[-\frac{1}{2}\operatorname{tr}(S\Sigma^{-1})\right]. \end{aligned}$$

Multiplying the prior with the likelihood yields the posterior:

$$\begin{aligned} f(\alpha,\Sigma|y) \propto f(y|\alpha,\Sigma)f(\alpha,\Sigma) \propto |\Sigma|^{-\frac{T+\underline{\nu}+M+1}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[(\underline{S}+S)\Sigma^{-1}\right]\right\} \\ \times \exp\left\{-\frac{1}{2}\underbrace{\left[(\alpha-\underline{\alpha})'(\Sigma^{-1}\otimes\underline{V}^{-1})(\alpha-\underline{\alpha})+(\alpha-\hat{\alpha})'(\Sigma^{-1}\otimes X'X)(\alpha-\hat{\alpha})\right]}_{\equiv\kappa}\right\}. \end{aligned} \tag{13}$$

Note that the argument κ in the second exponential function can be written as

$$\begin{aligned}
\kappa &= \alpha'(\Sigma^{-1} \otimes \underline{V}^{-1})\alpha - 2\underline{\alpha}'(\Sigma^{-1} \otimes \underline{V}^{-1})\alpha + \underline{\alpha}'(\Sigma^{-1} \otimes \underline{V}^{-1})\underline{\alpha} \\
&\quad + \alpha'(\Sigma^{-1} \otimes X'X)\alpha - 2\hat{\alpha}'(\Sigma^{-1} \otimes X'X)\alpha + \hat{\alpha}'(\Sigma^{-1} \otimes X'X)\hat{\alpha} \\
&= \alpha'[\Sigma^{-1} \otimes (\underline{V}^{-1} + X'X)]\alpha - 2[\underline{\alpha}'(\Sigma^{-1} \otimes \underline{V}^{-1}) + \hat{\alpha}'(\Sigma^{-1} \otimes X'X)]\alpha \\
&\quad + \underline{\alpha}'(\Sigma^{-1} \otimes \underline{V}^{-1})\underline{\alpha} + \hat{\alpha}'(\Sigma^{-1} \otimes X'X)\hat{\alpha} \\
&= \alpha'[\Sigma^{-1} \otimes (\underline{V}^{-1} + X'X)]\alpha - 2[\underline{\alpha}'(\Sigma^{-1} \otimes \underline{V}^{-1}) + \hat{\alpha}'(\Sigma^{-1} \otimes X'X)]\alpha \\
&\quad + \operatorname{tr}(\underline{A}'\underline{V}^{-1}\underline{A}\Sigma^{-1}) + \operatorname{tr}(\hat{A}'X'X\hat{A}\Sigma^{-1}), \qquad (14)
\end{aligned}$$

where $\operatorname{vec}(\underline{A}) = \underline{\alpha}$. Substituting into (13) and separating into two exponential functions (one that contains α and Σ and one that contains only Σ) yields

$$\begin{aligned}
f(\alpha, \Sigma|y) &\propto |\Sigma|^{-\frac{T+\underline{\nu}+Mp+M+2}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[(\underline{S} + S + \underline{A}'\underline{V}^{-1}\underline{A} + \hat{A}'X'X\hat{A})\Sigma^{-1}\right]\right\} \\
&\times \exp\left\{-\frac{1}{2}\left[\alpha'[\Sigma^{-1} \otimes (\underline{V}^{-1} + X'X)]\alpha - 2[\underline{\alpha}'(\Sigma^{-1} \otimes \underline{V}^{-1}) + \hat{\alpha}'(\Sigma^{-1} \otimes X'X)]\alpha\right]\right\}. \qquad (15)
\end{aligned}$$

Let us conjecture that (15) can be written as $f(\alpha, \Sigma|y) = f(\alpha|y, \Sigma)f(\Sigma|y)$, where $f(\alpha|y, \Sigma)$ is a normal distribution with posterior mean $\bar{\alpha}$, posterior variance matrix $\Sigma \otimes \bar{V}$ and pdf

$$\begin{aligned}
f(\alpha|y, \Sigma) &\propto |\Sigma \otimes \bar{V}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\alpha - \bar{\alpha})'(\Sigma^{-1} \otimes \bar{V}^{-1})(\alpha - \bar{\alpha})\right] \\
&\propto |\Sigma|^{-\frac{Mp+1}{2}} |\bar{V}|^{-\frac{M}{2}} \\
&\times \exp\left\{-\frac{1}{2}\left[\alpha'(\Sigma^{-1} \otimes \bar{V}^{-1})\alpha - 2\bar{\alpha}'(\Sigma^{-1} \otimes \bar{V}^{-1})\alpha + \bar{\alpha}'(\Sigma^{-1} \otimes \bar{V}^{-1})\bar{\alpha}\right]\right\}. \qquad (16)
\end{aligned}$$

Comparison of (15) and (16) shows that $\Sigma^{-1} \otimes \bar{V}^{-1} = \Sigma^{-1} \otimes (\underline{V}^{-1} + X'X)$ and thus

$$\bar{V} = (\underline{V}^{-1} + X'X)^{-1}.$$

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

$$\bar{\alpha} = (\Sigma \otimes \bar{V})\left[(\Sigma^{-1} \otimes \underline{V}^{-1})\underline{\alpha} + (\Sigma^{-1} \otimes X'X)\hat{\alpha}\right] = (I_M \otimes \bar{V}\underline{V}^{-1})\underline{\alpha} + (I_M \otimes \bar{V}X'X)\hat{\alpha}$$

or, equivalently,

$$\bar{A} = \bar{V}\underline{V}^{-1}\underline{A} + \bar{V}X'X\hat{A}.$$

In a final step, let us find

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

Substituting (15) and (16) yields

$$
\begin{aligned}
f(\Sigma|y) &\propto \frac{|\Sigma|^{-\frac{T+\underline{\nu}+Mp+M+2}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[(\underline{S}+S+\underline{A}'\underline{V}^{-1}\underline{A}+\hat{A}'X'X\hat{A})\Sigma^{-1}\right]\right\}}{|\Sigma|^{-\frac{Mp+1}{2}}} \\
&\times \frac{\exp\left\{-\frac{1}{2}\left[\alpha'[\Sigma^{-1}\otimes(\underline{V}^{-1}+X'X)]\alpha - 2[\underline{\alpha}'(\Sigma^{-1}\otimes\underline{V}^{-1})+\hat{\alpha}'(\Sigma^{-1}\otimes X'X)]\alpha\right]\right\}}{\exp\left\{-\frac{1}{2}\left[\alpha'(\Sigma^{-1}\otimes\bar{V}^{-1})\alpha - 2\bar{\alpha}'(\Sigma^{-1}\otimes\bar{V}^{-1})\alpha+\bar{\alpha}'(\Sigma^{-1}\otimes\bar{V}^{-1})\bar{\alpha}\right]\right\}} \\
&\propto \frac{|\Sigma|^{-\frac{T+\underline{\nu}+M+1}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[(\underline{S}+S+\underline{A}'\underline{V}^{-1}\underline{A}+\hat{A}'X'X\hat{A})\Sigma^{-1}\right]\right\}}{\exp\left\{-\frac{1}{2}\left[\bar{\alpha}'(\Sigma^{-1}\otimes\bar{V}^{-1})\bar{\alpha}\right]\right\}}. \qquad (17)
\end{aligned}
$$

By noting that $\bar{\alpha}'(\Sigma^{-1}\otimes\bar{V}^{-1})\bar{\alpha} = \operatorname{tr}(\bar{A}'\bar{V}^{-1}\bar{A}\Sigma^{-1})$, this can be simplified to

$$
\begin{aligned}
f(\Sigma|y) &\propto |\Sigma|^{-\frac{T+\underline{\nu}+M+1}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[(\underline{S}+S+\underline{A}'\underline{V}^{-1}\underline{A}+\hat{A}'X'X\hat{A}-\bar{A}'\bar{V}^{-1}\bar{A})\Sigma^{-1}\right]\right\} \\
&\propto |\Sigma|^{-\frac{\bar{\nu}+M+1}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[\bar{S}\Sigma^{-1}\right]\right\}, \qquad (18)
\end{aligned}
$$

which is the kernel of an inverse Wishart pdf with parameters

$$\bar{S} = \underline{S}+S+\underline{A}'\underline{V}^{-1}\underline{A}+\hat{A}'X'X\hat{A}-\bar{A}'\bar{V}^{-1}\bar{A}$$

and

$$\bar{\nu} = T+\underline{\nu}.$$

Hence, $\Sigma|y \sim iW(\bar{S},\bar{\nu})$.

3.3 Independent Normal-Wishart prior

The independent Normal-Wishart prior assumes that the variance matrix follows an inverse Wishart distribution, $\Sigma \sim iW(\underline{S},\underline{\nu})$, with pdf

$$f(\Sigma) \propto |\Sigma|^{-\frac{\underline{\nu}+M+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}(\underline{S}\Sigma^{-1})\right],$$

and the mean parameters follow a normal distribution, $\alpha \sim \mathcal{N}(\underline{\alpha},\underline{W})$ with pdf

$$f(\alpha) \propto |\underline{W}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\alpha-\underline{\alpha})'\underline{W}^{-1}(\alpha-\underline{\alpha})\right].$$

Hence the joint prior pdf is

$$f(\alpha,\Sigma)\propto|\Sigma|^{-\frac{\underline{\nu}+M+1}{2}}\exp\left[-\frac{1}{2}(\alpha-\underline{\alpha})'\underline{W}^{-1}(\alpha-\underline{\alpha})\right]\exp\left[-\frac{1}{2}\operatorname{tr}(\underline{S}\Sigma^{-1})\right].$$

Recall that the likelihood is

$$f(y|\alpha,\Sigma)\propto|\Sigma|^{-\frac{T}{2}}\exp\left[-\frac{1}{2}(\alpha-\hat{\alpha})'(\Sigma^{-1}\otimes X'X)(\alpha-\hat{\alpha})\right]\exp\left[-\frac{1}{2}\operatorname{tr}(S\Sigma^{-1})\right].$$

Multiplying the prior with the likelihood yields the posterior:

$$\begin{aligned}f(\alpha,\Sigma|y)\propto{}&|\Sigma|^{-\frac{T+\underline{\nu}+M+1}{2}}\exp\left\{-\frac{1}{2}\operatorname{tr}\left[(\underline{S}+S)\Sigma^{-1}\right]\right\}\\&\times\exp\left\{-\frac{1}{2}\underbrace{\left[(\alpha-\underline{\alpha})'\underline{W}^{-1}(\alpha-\underline{\alpha})+(\alpha-\hat{\alpha})'(\Sigma^{-1}\otimes X'X)(\alpha-\hat{\alpha})\right]}_{\equiv\kappa}\right\}.\end{aligned}\tag{19}$$

Note that the argument κ in the second exponential function can be written as

$$\begin{aligned}\kappa&=\alpha'\underline{W}^{-1}\alpha-2\underline{\alpha}'\underline{W}^{-1}\alpha+\underline{\alpha}'\underline{W}^{-1}\underline{\alpha}+\alpha'(\Sigma^{-1}\otimes X'X)\alpha-2\hat{\alpha}'(\Sigma^{-1}\otimes X'X)\alpha\\&\quad+\hat{\alpha}'(\Sigma^{-1}\otimes X'X)\hat{\alpha}\\&=\alpha'(\underline{W}^{-1}+\Sigma^{-1}\otimes X'X)\alpha-2[\underline{\alpha}'\underline{W}^{-1}+\hat{\alpha}'(\Sigma^{-1}\otimes X'X)]\alpha+\underline{\alpha}'\underline{W}^{-1}\underline{\alpha}\\&\quad+\hat{\alpha}'(\Sigma^{-1}\otimes X'X)\hat{\alpha}\\&=\alpha'(\underline{W}^{-1}+\Sigma^{-1}\otimes X'X)\alpha-2[\underline{\alpha}'\underline{W}^{-1}+\hat{\alpha}'(\Sigma^{-1}\otimes X'X)]\alpha+\underline{\alpha}'\underline{W}^{-1}\underline{\alpha}\\&\quad+\operatorname{tr}(\hat{A}'X'X\hat{A}\Sigma^{-1}).\end{aligned}\tag{20}$$

Substituting into (19) and separating into two exponential functions (one that contains α and Σ and one that contains only Σ) yields the joint postrior

$$\begin{aligned}f(\alpha,\Sigma|y)\propto{}&|\Sigma|^{-\frac{T+\underline{\nu}+M+1}{2}}\exp\left\{-\frac{1}{2}\operatorname{tr}\left[(\underline{S}+S+\hat{A}'X'X\hat{A})\Sigma^{-1}\right]\right\}\\&\times\exp\left\{-\frac{1}{2}\left[\alpha'(\underline{W}^{-1}+\Sigma^{-1}\otimes X'X)\alpha-2[\underline{\alpha}'\underline{W}^{-1}+\hat{\alpha}'(\Sigma^{-1}\otimes X'X)]\alpha\right]\right\}.\end{aligned}\tag{21}$$

Part 1: Finding $f(\alpha|y,\Sigma)$. Let us conjecture that (21) can be written as $f(\alpha,\Sigma|y)=f(\alpha|y,\Sigma)f(\Sigma|y)$, where $f(\alpha|y,\Sigma)$ is a normal distribution with posterior mean $\bar{\alpha}$, posterior variance matrix $\bar{W}$ and pdf

$$
\begin{aligned}
f(\alpha|y,\Sigma) &\propto |\bar{W}|^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\alpha-\bar{\alpha})'\bar{W}^{-1}(\alpha-\bar{\alpha})\right] \\
&\propto |\bar{W}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\alpha'\bar{W}^{-1}\alpha - 2\bar{\alpha}'\bar{W}^{-1}\alpha + \bar{\alpha}'\bar{W}^{-1}\bar{\alpha}\right]\right\}. \qquad (22)
\end{aligned}
$$

Comparison of (21) and (22) shows that $\bar{W}^{-1} = \underline{W}^{-1} + \Sigma^{-1} \otimes X'X$ and thus

$$
\bar{W} = (\underline{W}^{-1} + \Sigma^{-1} \otimes X'X)^{-1}.
$$

In addition, $\bar{\alpha}'\bar{W}^{-1} = \underline{\alpha}'\underline{W}^{-1} + \hat{\alpha}'(\Sigma^{-1} \otimes X'X)$ and thus

$$
\bar{\alpha} = \bar{W}\left[\underline{W}^{-1}\underline{\alpha} + (\Sigma^{-1} \otimes X'X)\hat{\alpha}\right].
$$

Next, let us find

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

Substituting (21) and (22) yields

$$
\begin{aligned}
f(\Sigma|y) &\propto \frac{|\Sigma|^{-\frac{T+\underline{\nu}+M+1}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[(\underline{S}+S+\hat{A}'X'X\hat{A})\Sigma^{-1}\right]\right\}}{|\bar{W}|^{-\frac{1}{2}}} \\
&\quad \times \frac{\exp\left\{-\frac{1}{2}\left[\alpha'(\underline{W}^{-1}+\Sigma^{-1}\otimes X'X)\alpha - 2[\underline{\alpha}'\underline{W}^{-1} + \hat{\alpha}'(\Sigma^{-1}\otimes X'X)]\alpha\right]\right\}}{\exp\left\{-\frac{1}{2}\left[\alpha'\bar{W}^{-1}\alpha - 2\bar{\alpha}'\bar{W}^{-1}\alpha + \bar{\alpha}'\bar{W}^{-1}\bar{\alpha}\right]\right\}} \\
&\propto \frac{|\Sigma|^{-\frac{T+\underline{\nu}+M+1}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[(\underline{S}+S+\hat{A}'X'X\hat{A})\Sigma^{-1}\right]\right\}}{|(\underline{W}^{-1}+\Sigma^{-1}\otimes X'X)^{-1}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left[\bar{\alpha}'(\underline{W}^{-1}+\Sigma^{-1}\otimes X'X)\bar{\alpha}\right]\right\}}. \qquad (23)
\end{aligned}
$$

Unfortunately, this is not the kernel of an inverse Wishart distribution (or any other known distribution).

Part 2: Finding $f(\Sigma|y,\alpha)$. Let us conjecture that (21) can be written as $f(\alpha,\Sigma|y) = f(\Sigma|y,\alpha)f(\alpha|y)$, where $f(\Sigma|y,\alpha)$ is an inverse Wishart distribution with posterior parameters $\bar{S}$ and $\bar{\nu}$ and pdf kernel

$$
f(\Sigma|y,\alpha) \propto |\Sigma|^{-\frac{\bar{\nu}+M+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}(\bar{S}\Sigma^{-1})\right]. \qquad (24)
$$

Using $(\alpha-\hat{\alpha})'(\Sigma^{-1}\otimes X'X)(\alpha-\hat{\alpha}) = \operatorname{tr}[(A-\hat{A})'X'X(A-\hat{A})\Sigma^{-1}]$, re-write the joint posterior (19) accordingly:

$$\begin{aligned}
f(\alpha, \Sigma|y) \propto |\Sigma|^{-\frac{T+\underline{\nu}+M+1}{2}} & \exp\left\{-\frac{1}{2}\operatorname{tr}\left[(\underline{S}+S)\Sigma^{-1}\right]\right\} \\
& \times \exp\left\{-\frac{1}{2}\left[(\alpha-\underline{\alpha})'\underline{W}^{-1}(\alpha-\underline{\alpha})+(\alpha-\hat{\alpha})'(\Sigma^{-1}\otimes X'X)(\alpha-\hat{\alpha})\right]\right\} \\
\propto & |\Sigma|^{-\frac{T+\underline{\nu}+M+1}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[(\underline{S}+S+(A-\hat{A})'X'X(A-\hat{A}))\Sigma^{-1}\right]\right\} \\
& \times \exp\left\{-\frac{1}{2}\left[(\alpha-\underline{\alpha})'\underline{W}^{-1}(\alpha-\underline{\alpha})\right]\right\}.
\end{aligned} \tag{25}$$

Comparison of (25) and (24) shows that

$$\bar{\nu} = T + \underline{\nu}$$

and

$$\bar{S} = \underline{S} + S + (A-\hat{A})'X'X(A-\hat{A}).$$

Recalling that $S = Y'M_X Y$, this can equivalently be expressed as

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

Hence, the posterior pdf kernel for Σ conditional on α that includes all parts conainting α is

$$f(\Sigma|y,\alpha) \propto |\bar{S}|^{\frac{\bar{\nu}}{2}}|\Sigma|^{-\frac{\bar{\nu}+M+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}(\bar{S}\Sigma^{-1})\right] \tag{26}$$

because $\bar{S}$ is a function of α. Now, let us find

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

Substituting (25) and (26) yields

$$
\begin{aligned}
f(\alpha|y) &\propto \frac{|\Sigma|^{-\frac{T+\underline{\nu}+M+1}{2}} \exp\left\{-\frac{1}{2}\operatorname{tr}\left[(\underline{S}+S+(A-\hat{A})'X'X(A-\hat{A}))\Sigma^{-1}\right]\right\}}{|\Sigma|^{-\frac{\bar{\nu}+M+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}(\bar{S}\Sigma^{-1})\right]} \\
&\quad \times \frac{\exp\left\{-\frac{1}{2}\left[(\alpha-\underline{\alpha})'\underline{W}^{-1}(\alpha-\underline{\alpha})\right]\right\}}{|\bar{S}|^{\frac{\bar{\nu}}{2}}} \\
&\propto |\underline{S}+(Y-XA)'(Y-XA)|^{-\frac{\bar{\nu}}{2}} \exp\left\{-\frac{1}{2}\left[(\alpha-\underline{\alpha})'\underline{W}^{-1}(\alpha-\underline{\alpha})\right]\right\}.
\end{aligned}
$$

Unfortunately, this is not the kernel of a normal distribution (or any other known distribution).

Appendix: 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 V and ν, $H \sim W(V, \nu)$. Then it has pdf

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

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 B is an $M \times M$ symmetric and positive definite scale matrix. The expectation is

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

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

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

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

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 $H \sim W(V, \nu)$. Then $\Sigma \equiv H^{-1} \sim iW(V^{-1}, \nu)$.