Matrix T derivation

October 25, 2013

1 Joint-conditional method

Let's start with the density of Y, conditional on Σ.

$$
\begin{aligned}
P(Y|\Sigma,\cdot) &= \prod_{t=1}^{T} (2\pi)^{-\frac{NJ}{2}} |Q_t|^{-\frac{J}{2}} |\Sigma|^{-\frac{N}{2}} \exp\left[-\frac{1}{2}\,\mathrm{tr}\left[\left(Y_t - f_t\right)' Q_t^{-1} \left(Y_t - f_t\right)\Sigma^{-1}\right]\right] && (1) \quad \text{eq:py} \\
&= (2\pi)^{-\frac{NJT}{2}} \left(\prod_{t=1}^{T} |Q_t|^{-\frac{J}{2}}\right) |\Sigma|^{-\frac{NT}{2}} \exp\left[-\frac{1}{2}\,\mathrm{tr}\left[\sum_{t=1}^{T}\left(Y_t - f_t\right)' Q_t^{-1} \left(Y_t - f_t\right)\Sigma^{-1}\right]\right] && (2) \\
&= \mathcal{K}|\Sigma|^{-\frac{NT}{2}} \exp\left[-\frac{1}{2}\,\mathrm{tr}\left(\mathcal{A}\Sigma^{-1}\right)\right] && (3)
\end{aligned}
$$

where

$$
\mathcal{K} = (2\pi)^{-\frac{NJT}{2}} \left(\prod_{t=1}^{T} |Q_t|^{-\frac{J}{2}}\right) \tag{4}
$$

eq:2

and

$$
\mathcal{A} = \sum_{t=1}^{T} \left(Y_t - f_t\right)' Q_t^{-1} \left(Y_t - f_t\right) \tag{5}
$$

eq:defA

Next, we give Σ an inverse-Wishart prior.

$$
P\left(\Sigma|\cdot\right) = \mathcal{M}|\Sigma|^{-\frac{\nu_0 - J + 1}{2}} \exp\left[-\frac{1}{2}\,\mathrm{tr}\left(\Omega_0\Sigma^{-1}\right)\right] \tag{6}
$$

eq:pSig

where

$$
\mathcal{M} = \frac{|\Omega_0|^{\frac{\nu_0}{2}}}{2^{\frac{J\nu_0}{2}}\Gamma_J\left(\frac{\nu_0}{2}\right)} \tag{7}
$$

eq:defM

The $\Gamma_J(\cdot)$ term is the multivariate gamma function.
Combining terms, and setting up the integral, we get

$$
P(Y|\cdot) = \mathcal{K}\mathcal{M} \int |\Sigma|^{-\frac{NT + \nu_0 - J + 1}{2}} \exp\left[-\frac{1}{2}\,\mathrm{tr}\left(\left(\Omega_0 + \mathcal{A}\right)\Sigma^{-1}\right)\right] d\Sigma \tag{8}
$$

eq:3

Notice that the integrand looks a lot like an inverse-Wishart density itself, but with $NT + \nu_0$ degrees of freedom and scale parameter $(\Omega_0 + \mathcal{A})$. Since integrals of densities over the entire domain of the variable of interest must equal 1, we know that

$$\int |\Sigma|^{-\frac{NT+\nu_0-J+1}{2}} \exp\left[-\frac{1}{2}\operatorname{tr}\left((\Omega_0+\mathcal{A})\,\Sigma^{-1}\right)\right] d\Sigma = \frac{2^{\frac{(NT+\nu_0)J}{2}}\Gamma_J\left(\frac{NT+\nu_0}{2}\right)}{|\Omega_0+\mathcal{A}|^{\frac{NT+\nu_0}{2}}} \tag{9}$$

eq:1

(The right side of that equation is the inverse of the normalizing constant for the inverse Wishart).

Therefore, by substitution, we get

$$P(Y|\cdot) = \mathcal{K}\mathcal{M}\left(\frac{2^{\frac{(NT+\nu_0)J}{2}}\Gamma_J\left(\frac{NT+\nu_0}{2}\right)}{|\Omega_0+\mathcal{A}|^{\frac{NT+\nu_0}{2}}}\right) \tag{10}$$

eq:4

$$= \left(\prod_{t=1}^{T}|Q_t|^{-\frac{J}{2}}\right)\pi^{-\frac{NTJ}{2}}\frac{\Gamma_J\left(\frac{NT+\nu_0}{2}\right)}{\Gamma_J\left(\frac{\nu_0}{2}\right)}\frac{|\Omega_0|^{\frac{\nu_0}{2}}}{|\Omega_0+\sum_{t=1}^{T}(Y_t-f_t)'\,Q_t^{-1}\,(Y_t-f_t)\,|^{\frac{NT+\nu_0}{2}}} \tag{11}$$

2 Bayes updating method

The data likelihood at time t, conditional on all past data, is

$$P(Y_t|Y_{1:t-1},\Sigma,\cdot) = (2\pi)^{-\frac{NJ}{2}}\,|Q_t|^{-\frac{J}{2}}|\Sigma|^{-\frac{N}{2}}\exp\left[-\frac{1}{2}\operatorname{tr}\left[(Y_t-f_t)'\,Q_t^{-1}\,(Y_t-f_t)\,\Sigma^{-1}\right]\right] \tag{12}$$

eq:5

Suppose we place an inverse-Wishart prior on Σ, with ν_0 degrees of freedom and scale parameter Ω_0. Since the matrix normal and inverse Wishart densities form a conjugate pair, the posterior density of Σ is also inverse Wishart, with updates:

$$\nu_t \leftarrow \nu_{t-1} + N \tag{13}$$

eq:6

$$\Omega_t \leftarrow \Omega_{t-1} + a_t \tag{14}$$

where $a_t = (Y_t - f_t)'\,Q_t^{-1}\,(Y_t - f_t)$.

Therefore, for a single period t, we can integrate out Σ by integrating over the posterior density of Σ, conditional on accumulated knowledge through time $t-1$. This involves integrating over an inverse Wishart(ν_{t-1},Ω_{t-1}) density.

$$P(Y_t|Y_{1:t-1},\cdot) = (2\pi)^{-\frac{NJ}{2}}\,|Q_t|^{-\frac{J}{2}}\frac{|\Omega_{t-1}|^{\frac{\nu_{t-1}}{2}}}{2^{\frac{J\nu_{t-1}}{2}}\Gamma_J\left(\frac{\nu_{t-1}}{2}\right)}\int|\Sigma|^{-\frac{N+\nu_{t-1}+J+1}{2}}\exp\left[-\frac{1}{2}\operatorname{tr}\left[(\Omega_{t-1}+a_t)\,\Sigma^{-1}\right]\right]d\Sigma \tag{15}$$

eq:7

Given the updates for ν_t and Ω_t, we can rewrite this as

$$P(Y_t|Y_{1:t-1},\cdot) = (2\pi)^{-\frac{NJ}{2}}\,|Q_t|^{-\frac{J}{2}}\frac{|\Omega_{t-1}|^{\frac{\nu_{t-1}}{2}}}{2^{\frac{J\nu_{t-1}}{2}}\Gamma_J\left(\frac{\nu_{t-1}}{2}\right)}\int|\Sigma|^{-\frac{\nu_t+J+1}{2}}\exp\left[-\frac{1}{2}\operatorname{tr}\left[\Omega_t\Sigma^{-1}\right]\right]d\Sigma \tag{16}$$

eq:8

$$= \pi^{-\frac{NJ}{2}}|Q_t|^{-\frac{J}{2}}\frac{\Gamma_J\left(\frac{\nu_t}{2}\right)}{\Gamma_J\left(\frac{\nu_{t-1}}{2}\right)}\frac{|\Omega_{t-1}|^{\frac{\nu_{t-1}}{2}}}{|\Omega_t|^{\frac{\nu_t}{2}}} \tag{17}$$

To get the full data likelihood, we note that a $t-1$ term in one period is the t term in the next period, so the only ν and Ω terms that do not get canceled out are the prior parameters ν_0 and Ω_0, and the posterior parameters after the last observed period. Those posteriors are

$$\nu_T \leftarrow \nu_0 + TN \tag{18}$$

eq:10

$$\Omega_T \leftarrow \Omega_0 + \sum_{t=1}^{T} a_t = \Omega_0 + \mathcal{A} \tag{19}$$

Thus, we get the following data likelihood

$$P(Y|\cdot) = \prod_{t=1}^{T} P(Y_t|y_{1:t-1},\cdot) \tag{20}$$

eq:9

$$= \pi^{-\frac{NJT}{2}} \left(\prod_{t=1}^{T} |Q_t|^{-\frac{J}{2}}\right) \frac{\Gamma_J\left(\frac{\nu_0+TN}{2}\right)}{\Gamma_J\left(\frac{\nu_0}{2}\right)} \frac{|\Omega_0|^{-\frac{\nu_0}{2}}}{|\Omega_0 + \sum_{t=1}^{T}(Y_t - f_t)' Q_t^{-1} (Y_t - f_t)|^{\frac{\nu_0+TN}{2}}} \tag{21}$$

which is exactly what we got when we used the joint-conditional method.

3 Matrix Normal Bayesian Estimation of linear parameters

Let's suppose we have a model (conditional on V_1 and Σ and on state variables Θ_{11t}):

$$Y_t^* = Y_t - F_{11t}\Theta_{11t} = F_{12t}\Theta_{12} + v_{1t}$$

with Y_t being $N \times J$, F_{12t} is a covariate matrix (dimension $N \times K$). The term v_{1t} has a matrix normal distribution $N(0, V_1, \Sigma)$, assumed known. We put a prior on Θ_{12} as $N(\underline{\Theta}, \underline{S}, \Sigma)$. With $\underline{S}$ being $K \times K$, V_1 is $N \times N$, $\underline{\Theta}$ is $K \times J$, as is Θ_{12}. Finally, Σ is $J \times J$.

This means the posterior distribution is:

$$\begin{aligned} P(\Theta|\Sigma, Y_t^*, F_{12t}, V_1) = & \quad (2\pi)^{-\frac{NJ}{2}} |V_1|^{-\frac{J}{2}} |\Sigma|^{-\frac{N}{2}} (2\pi)^{-\frac{NK}{2}} |\underline{S}|^{\frac{K}{2}} |\Sigma|^{-\frac{N}{2}} \\ & \exp\left(-tr\frac{1}{2}\left[\left\{(Y_t^* - F_{12t}\Theta_{12})' V^{-1} (Y_t^* - F_{12t}\Theta_{12}) + (\Theta_{12} - \underline{\Theta}_{12})' \underline{S}^{-1} (\Theta_{12} - \underline{\Theta}_{12})\Sigma^{-1}\right\}\right]\right) \\ = & \quad (2\pi)^{-\frac{N(J+K)}{2}} |V_1|^{-\frac{J}{2}} |\Sigma|^{-\frac{2N}{2}} |\underline{S}|^{\frac{K}{2}} \\ & \exp\left(-tr\frac{1}{2}\left[(Y_t^* - F_{12t}\Theta_{12})' V^{-1} (Y_t^* - F_{12t}\Theta_{12}) + (\Theta_{12} - \underline{\Theta}_{12})' \underline{S}^{-1} (\Theta_{12} - \underline{\Theta}_{12})\Sigma^{-1}\right]\right) \end{aligned} \tag{22}$$

Focusing on the term in the exponential trace, before the post-multiplication of the Σ^{-1} term (both quadratics are dimension $J \times J$ so are conformable for the trace):

$$= (Y_t^* - F_{12t}\Theta_{12})' V^{-1} (Y_t^* - F_{12t}\Theta_{12}) + (\Theta_{12} - \underline{\Theta}_{12})' \underline{S}^{-1} (\Theta_{12} - \underline{\Theta}_{12}) \tag{23}$$

expanding out the quadratic terms:

$$\begin{aligned} = & Y_t^{*\prime} V^{-1} Y_t^* - (F_{12t}\Theta_{12})' V^{-1} Y_t^* - Y_t^* V^{-1} F_{12t}\Theta_{12} + (F_{12t}\Theta_{12})' V^{-1} F_{12t}\Theta_{12} \\ & + \Theta_{12}' \underline{S}^{-1} \Theta_{12} - \underline{\Theta}_{12}' \underline{S}^{-1} \Theta_{12} - \Theta_{12}' \underline{S}^{-1} \underline{\Theta}_{12} + \underline{\Theta}_{12}' \underline{S}^{-1} \underline{\Theta}_{12} \end{aligned} \tag{24}$$

dropping terms that do not depend on Θ_{12}:

$$\begin{aligned} = & (F_{12t}\Theta_{12})' V^{-1} Y_t^* - Y_t^* V^{-1} F_{12t}\Theta_{12} + (F_{12t}\Theta_{12})' V^{-1} F_{12t}\Theta_{12} \\ & + \Theta_{12}' \underline{S}^{-1} \Theta_{12} - \underline{\Theta}_{12}' \underline{S}^{-1} \Theta_{12} - \Theta_{12}' \underline{S}^{-1} \underline{\Theta}_{12} \end{aligned} \tag{25}$$

which can be rewritten:

$$\Theta_{12}' \left[F_{12t}' V^{-1} F_{12t} + \underline{S}^{-1} \right] \Theta_{12} \quad - \Theta_{12}' \left[F_{12t}' V^{-1} Y_t^* - \underline{S}^{-1} \underline{\Theta}_{12} \right] - \left[Y_t^{*\prime} V^{-1} F_{12t} + \underline{\Theta}_{12}' \underline{S}^{-1} \right] \Theta_{12} \tag{26}$$

Then for the magic bit. We write $\mathbf{A} = F_{12t}' V^{-1} F_{12t} + \underline{S}^{-1}$, and $\mathbf{b} = F_{12t}' V^{-1} Y_t^* - \underline{S}^{-1} \underline{\Theta}_{12}$, and recognizing that V and $\underline{S}$ are symmetric. Then we have:

$$\Theta_{12}' \mathbf{A} \Theta_{12} - \Theta_{12}' \mathbf{b} - \mathbf{b}' \Theta_{12} \tag{27}$$

We can complete the square by adding a term that does not depend on Θ_{12}, we use: $\mathbf{b}' \mathbf{A}^{-1} \mathbf{b}$.

$$\Theta_{12}' \mathbf{A} \Theta_{12} - \Theta_{12}' \mathbf{b} - \mathbf{b}' \Theta_{12} + \mathbf{b}' \mathbf{A}^{-1} \mathbf{b} \tag{28}$$

Noting that A is symmetric and invertible (it is the weighted sum of two symmetric and full rank matrixes), the identity matrix $\mathbf{I} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1}$ then we can rewrite the above as:

$$\Theta_{12}' \mathbf{A} \Theta_{12} - \Theta_{12}' \mathbf{A}\mathbf{A}^{-1} \mathbf{b} - \mathbf{b}' \mathbf{A}^{-1} \mathbf{A} \Theta_{12} + \mathbf{b}' \mathbf{A}^{-1} \mathbf{A} \mathbf{A}^{-1} \mathbf{b} \tag{29}$$

Now let us denote $\hat{\Theta}_{12} = \mathbf{A}^{-1}\mathbf{b}$ and $\hat{S} = \mathbf{A}^{-1}$. Then we can write:

$$\Theta_{12}' \hat{S}^{-1} \Theta_{12} - \Theta_{12}' \hat{S}^{-1} \hat{\Theta}_{12} - \hat{\Theta}_{12} \hat{S}^{-1} \Theta_{12} + \hat{\Theta}_{12}' \hat{S}^{-1} \hat{\Theta}_{12} \tag{30}$$

which we can factor to rewrite:

$$(\Theta_{12} - \hat{\Theta}_{12})' \hat{S}^{-1} (\Theta_{12} - \hat{\Theta}_{12}) \tag{31}$$

Substituting back in the trace above in eqn:ft (22), we therefore have that the posterior conditional for Θ_{12} is $N(\hat{\Theta}_{12}, \hat{S}, \Sigma)$. The full expression for each moment is:

$$\begin{aligned} \hat{\Theta}_{12} &= \mathbf{A}^{-1}\mathbf{b} = \left[F_{12t}' V^{-1} F_{12t} + \underline{S}^{-1} \right]^{-1} F_{12t}' V^{-1} Y_t^* - \underline{S}^{-1} \underline{\Theta}_{12} && (32) \\ \hat{S} &= \mathbf{A}^{-1} = \left[F_{12t}' V^{-1} F_{12t} + \underline{S}^{-1} \right]^{-1} && (33) \end{aligned}$$

This was for just one Y_t^*. Let's stack all of them together, so that $Y^* = \left[Y_1^* \vdots Y_t^* \vdots \ldots \vdots Y_T^* \right]$, and same with F_{12}, so that dimension of Y^* is $NT \times J$ and of F_{12} is $NT \times K$. The matrix Θ_{12} is still the same. Now the variance matrix can be written as $\tilde{V}_1 = I_T \otimes V_1$, being dimension $NT \times NT$. For the likelihood, we have a matrix normal so that $Y^* \sim N(F_{12}\Theta_{12}, \tilde{V}_1, \Sigma)$. The above logic still applies as above. I wonder if we can rewrite this to avoid having to calculate the large matrix $\tilde{V}$.

$$\begin{aligned} P(\Theta|\Sigma, Y_t^*, F_{12t}, V_1) &= (2\pi)^{-\frac{NK}{2}} |\underline{S}|^{\frac{K}{2}} |\Sigma|^{-\frac{N}{2}} \exp\left(-tr\frac{1}{2} \left[(\Theta_{12} - \underline{\Theta}_{12})' \underline{S}^{-1} (\Theta_{12} - \underline{\Theta}_{12}) \Sigma^{-1} \right] \right) \\ &\times \prod_{t=1}^{T} (2\pi)^{-\frac{NJ}{2}} |V_1|^{-\frac{J}{2}} |\Sigma|^{-\frac{N}{2}} \exp\left(-tr\frac{1}{2} \left[\left\{ (Y_t^* - F_{12t}\Theta_{12})' V^{-1} (Y_t^* - F_{12t}\Theta_{12}) \Sigma^{-1} \right\} \right] \right) \end{aligned} \tag{34}$$

So pulling the product in (and focusing on the second part of the right hand side):

$$\begin{aligned} P(\Theta|\Sigma, Y_t^*, F_{12t}, V_1) &= (2\pi)^{-\frac{NK}{2}} |\underline{S}|^{\frac{K}{2}} |\Sigma|^{-\frac{N}{2}} \exp\left(-tr\frac{1}{2}\left[(\Theta_{12}-\underline{\Theta}_{12})'\underline{S}^{-1}(\Theta_{12}-\underline{\Theta}_{12})\Sigma^{-1}\right]\right) \\ &\times (2\pi)^{-\frac{TNJ}{2}} |V_1|^{-\frac{TJ}{2}} |\Sigma|^{-\frac{TN}{2}} \exp\left(-tr\frac{1}{2}\left[\sum_{t=1}^{T}\left\{(Y_t^*-F_{12t}\Theta_{12})'V^{-1}(Y_t^*-F_{12t}\Theta_{12})\Sigma^{-1}\right\}\right]\right) \end{aligned} \tag{35}$$

Focusing on the terms in the exponential, expanding the quadratic (and dropping the terms not depending on Θ_{12}, we get:

$$\begin{aligned} = \sum_{t=1}^{T}(F_{12t}\Theta_{12})'V^{-1}Y_t^* - \sum_{t=1}^{T} Y_t^* V^{-1} F_{12t}\Theta_{12} + \sum_{t=1}^{T}(F_{12t}\Theta_{12})'V^{-1}F_{12t}\Theta_{12} \\ +\Theta_{12}'\underline{S}^{-1}\Theta_{12} - \underline{\Theta}_{12}'\underline{S}^{-1}\Theta_{12} - \Theta_{12}'\underline{S}^{-1}\underline{\Theta}_{12} \end{aligned} \tag{36}$$

which can be rewritten:

$$\Theta_{12}'\left[\sum_{t=1}^{T}\left(F_{12t}'V^{-1}F_{12t}\right)+\underline{S}^{-1}\right]\Theta_{12} \;-\Theta_{12}'\left[\sum_{t=1}^{T}\left(F_{12t}'V^{-1}Y_t^*\right)-\underline{S}^{-1}\underline{\Theta}_{12}\right] - \left[\sum_{t=1}^{T}\left(Y_t^{*\prime}V^{-1}F_{12t}\right)+\underline{\Theta}_{12}'\underline{S}^{-1}\right]\Theta_{12} \tag{37}$$

so using the trick from earlier, we write $\mathbf{A} = \sum_{t=1}^{T}\left(F_{12t}'V^{-1}F_{12t}\right)+\underline{S}^{-1}$, and $\mathbf{b} = \sum_{t=1}^{T}\left(F_{12t}'V^{-1}Y_t^*\right) - \underline{S}^{-1}\underline{\Theta}_{12}$, then following the logic from above we have:

$$\begin{aligned} \hat{\Theta}_{12} &= \mathbf{A}^{-1}\mathbf{b} = \left[\sum_{t=1}^{T}\left(F_{12t}'V^{-1}F_{12t}\right)+\underline{S}^{-1}\right]^{-1}\sum_{t=1}^{T}\left(F_{12t}'V^{-1}Y_t^*\right)-\underline{S}^{-1}\underline{\Theta}_{12} & (38) \\ \hat{S} &= \mathbf{A}^{-1} = \left[\sum_{t=1}^{T}\left(F_{12t}'V^{-1}F_{12t}\right)+\underline{S}^{-1}\right]^{-1} & (39) \end{aligned}$$